Google Data Analytics Certificate: Capstone Project

“Visualizing Financial Statements”

By

Nate Boyle

I. Introduction and Purpose.

This Data Analysis project is for the Google Data Analytics Certificate Program Capstone: Complete a Case Study. My name is Nate Boyle, I am 35 years old, originally from California, and now residing in Las Vegas, Nevada. I hold two associate degrees from Allan Hancock Community College in Santa Maria, CA, one in Computer Science and one in Mathematics, one bachelor's degree from the University of California, Davis in Economics with a specialization in Data Analytics and Economics Analysis, and lastly, a Master of Accounting from Cal State East Bay in Hayward, CA.

Using the Google Cloud Marketplace, I began searching for publicly available datasets. Given my professional experience as an auditor of financial statements at public accounting firms and my academic history mentioned above, I was specifically on the lookout for datasets involving financial information. Luckily, I found the **SEC Public Dataset**, titled "sec_quarterly_financials" in BigQuery, Google's data warehouse . The dataset is comprised of data from various forms submitted to the SEC, which was right in line with something I was looking for. Below is a screenshot of the front page of the dataset with a brief description of it with additional information below that.

SEC

SEC Public Dataset

U.S. Securities and Exchange Commission

U.S. Securities and Exchange Commission (SEC) financial data

VIEW DATASET

OVERVIEW SAMPLES

Overview

In the U.S. public companies, certain insiders and broker-dealers are required to regularly file with the SEC. The SEC makes this data available online for anybody to view and use via their Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database. The SEC updates this data every quarter going back to January, 2009. To aid analysis a quick summary view of the data has been created that is not available in the original dataset. The quick summary view pulls together signals into a single table that otherwise would have to be joined from multiple tables and enables a more streamlined user experience. This public dataset is hosted in Google BigQuery and is included in BigQuery's 1TB/mo of free tier processing. This means that each user receives 1TB of free BigQuery processing every month, which can be used to run queries on this public dataset. Watch this short video to learn how to get started quickly using BigQuery to access public datasets.

Dataset ID	bigquery-public-data.sec_quarterly_financials
Created	Sep 12, 2017, 7:59:04 AM UTC-7
Default table expiration	Never
Last modified	Sep 20, 2022, 12:58:20 AM UTC-7
Data location	US
Description	In the U.S. public companies, certain insiders and broker-dealers are required to regularly file with the SEC. The SEC makes this data available online for anybody to view and use via their Electronic Data Gathering, Analysis, and Retrieval (EDGAR) database. The SEC updates this data every quarter going back to January, 2009. For more information please see this site. To aid analysis a quick summary view of the data has been created that is not available in the original dataset. The quick summary view pulls together signals into a single table that otherwise would have to be joined from multiple tables and enables a more streamlined user experience.

Now that I had found my data set I was ready to begin the next part of the project.

II. Using SQL and Excel to transform data.

Most public datasets within the Google Cloud Datasets Marketplace come with multiple tables linked by keys, so the next task was determining which tables within the SEC dataset were relevant to the analyses that were to be performed. Below is a screenshot of the tables found within the sec_quarterly_financials dataset.

- sec_quarterly_financials
  - calculation
  - dimension
  - measure_tag
  - numbers
  - presentation
  - quick_summary
  - rendering
  - sic_codes
  - submission
  - txt

After searching through each table, the *quick_summary* table was determined to be the best to use, as it contained dollar amounts per account for each fiscal period, with fields for *company_name*, *measure_tag* (financial line item), *value* (dollar/share amount), and various other fields that would be necessary for queries. Below is a screen shot of each field that would be used from the *quick_summary* table.

Field name	Type	Mode	Collation	Default Value	Policy Tags	Description
company_name	STRING	NULLABLE				Name of registrant from the submission table. This corresponds to the name of the legal entity as recorded in EDGAR as of the filing date.
measure_tag	STRING	REQUIRED				The unique identifier (name) for a tag in a specific taxonomy release. measure_tag: [tag]
period_end_date	STRING	REQUIRED				The end date for the data value, rounded to the nearest month end. period_end_date: [ddate]
value	FLOAT	NULLABLE				The value. This is not scaled, it is as found in the Interactive Data file, but is rounded to four digits to the right of the decimal point.
number_of_quarters	INTEGER	REQUIRED				The count of the number of quarters represented by the data value, rounded to the nearest whole number. \"0\" indicates it is a point-in-time value. number_of_quarters: [qtrs]
sic	STRING	NULLABLE				Standard Industrial Classification (SIC). Four digit code assigned by the Commission as of the filing date, indicating the registrant's type of business.
fiscal_year_end	STRING	NULLABLE				Fiscal Year End Date. fiscal_year_end: [fye]
form	STRING	NULLABLE				The submission type of the registrant's filing.
fiscal_year	INTEGER	NULLABLE				Fiscal Year Focus (as defined in EFM Ch. 6). fiscal_year: [fy]

Additionally, the *industry_title* field from the *sic_codes* table shown below would also end up being used.

Field name	Type	Mode	Collation	Default Value	Policy Tags	Description
sic_code	STRING	NULLABLE				
ad_office	STRING	NULLABLE				
industry_title	STRING	NULLABLE				

Once the tables from the dataset had been picked, querying was initiated for companies that were of interest to be included in the analysis. As a side note, BigQuery uses a form of SQL (Structured Query Language) to search through the tables and datasets stored in the data warehouse. In particular, BigQuery supports the "GoogleSQL" dialect of SQL.

Boeing was the company first queried for. Because the exact string *company_name* stored within the SEC dataset was unknown, the **LIKE** predicate was used to narrow down the company being looking for. It was also important to make sure that any company selected had a suitable number of fiscal years so trends analyzed over time, to accomplish this the *fiscal_year* field was included in the query, using the **ORDER BY** clause to have the years in ascending order.

```
SELECT DISTINCT company_name, fiscal_year
FROM `bigquery-public-data.sec_quarterly_financials.quick_summary` AS quick_summary
WHERE company_name LIKE "%BOEING%"
ORDER BY fiscal_year
```

Query results

JOB INFORMATION RESULTS JSON EXECUTION DETAILS

w	company_name	fiscal_year
4	BOEING CAPITAL CORP	2011
5	BOEING CO	2012
6	BOEING CAPITAL CORP	2012

When the *company_name* string had been found for a company that was being queried for, the search was narrowed down to only include rows that came from a company's 10-K. A Form 10-K is an annual report required by the U.S. Securities and Exchange Commission (SEC), that gives a comprehensive summary of a company's financial performance. The reason only 10-K was chosen was partially because annual data tends to be the most pertinent, especially since the project would be comparing balance sheet accounts to revenue, an income statement account, which is reset at the beginning of each fiscal year, but also because this would reduce the amount of time needed to be spent on the SEC EDGAR (Electronic Data Gathering, Analysis, and Retrieval system) website, because this way only one type of SEC form would have to be agreed to.

Through a combination of googling for publicly traded companies as well as seeing what companies had workable data within the dataset, thirty companies across ten industries were chosen.

Company	Industry
Boeing	Aerospace
Lockheed Martin	Aerospace
Northrup Grumman	Aerospace
Ford	Automotive
General Motors	Automotive
Tesla	Automotive
Bank of America	Banking
JP Morgan Chase	Banking
Citigroup	Banking
Aecom	Construction
Vulcan	Construction
Fluor	Construction
Disney	Entertainment
Time Warner	Entertainment
Netflix	Entertainment
Kraft	Food and Beverage
Coca-Cola	Food and Beverage
Pepsi	Food and Beverage
Google	Information Technology
Apple	Information Technology
Microsoft	Information Technology
Pfizer	Biotech
Amgen	Biotech
Thermo Fisher	Biotech
Walmart	Retail
Kroger	Retail
Amazon	Retail
Nike	Sporting Goods
Columbia	Sporting Goods
Under Armour	Sporting Goods

Before narrowing down the query for 10-Ks the tables would also include data for 10-Qs, quarterly reports.

```sql
SELECT DISTINCT company_name, fiscal_year, form
FROM `bigquery-public-data.sec_quarterly_financials.quick_summary` AS quick_summary
WHERE company_name = "BOEING CO"
ORDER BY fiscal_year

```

Query results

JOB INFORMATION RESULTS JSON EXECUTION DETAILS

w	company_name	fiscal_year	form
13	BOEING CO	2015	10-Q
14	BOEING CO	2015	10-K
15	BOEING CO	2016	10-Q
16	BOEING CO	2016	10-K

After using an **AND** conditional to set the *form* to "10-K" one would just get data from annual reports:

```sql
SELECT DISTINCT company_name, fiscal_year, form
FROM `bigquery-public-data.sec_quarterly_financials.quick_summary` AS quick_summary
WHERE company_name = "BOEING CO"
AND form = "10-K"
ORDER BY fiscal_year

```

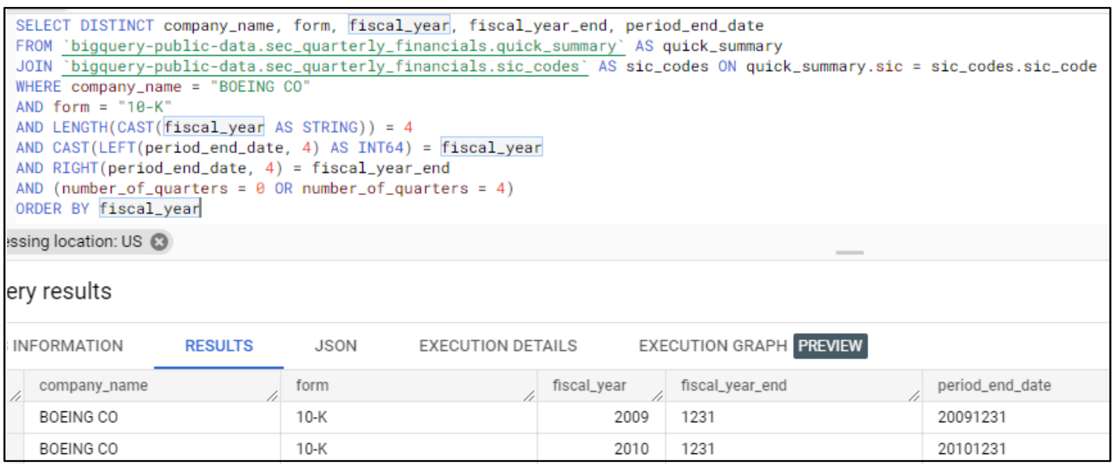
Query results

JOB INFORMATION RESULTS JSON EXECUTION DETAILS

ow	company_name	fiscal_year	form
1	BOEING CO	2009	10-K
2	BOEING CO	2010	10-K

If one has ever actually looked at a 10-K, one would know it often contains data from multiple fiscal periods and years, typically for the sake of comparing to current year/period results.

As stated previously, it was decided to only have year-end data specific to the fiscal year the 10-K was filed for, and no prior year or quarter data. To accomplish this various **AND** conditions were added to the query. As can be seen below the *period_end_date* field uses a "yyyymmdd" date format, so making sure both the "yyyy" and "mmdd" portions were correct was necessary to have only year end data of the fiscal year being filed. First, the **LENGTH** of the *fiscal_year* field had to be set to 4 characters in length, a few datasets had *fiscal_year* values that erroneously didn't. Then, to ensure that the periods selected were only for the fiscal year of filing, the query used the **LEFT** function to capture just the year ("yyyy") portion of the *period_end_date* field, and then used the **=** operator so that it matched the *fiscal_year* field. Then, similarly to the previous condition, the query used the **RIGHT** function to make sure that the month and day ("mmdd") portion of the *period_end_date* field matched the *fiscal_year_end* in terms of the quarter, and not just the year. Finally, to make sure the query was getting income statement data from the whole year and not just the fourth quarter, as both periods have the same *period_end_date,* the query needed to have a condition that set the *number_of_quarters* to **4**, but also had to set it for **0** for balance sheet data, because as mentioned above, balance sheet data is cumulative over time, and not just quarterly, so the dataset had **0**s for balance sheet rows in the *number_of_quarters* field.



```
SELECT DISTINCT company_name, form, fiscal_year, fiscal_year_end, period_end_date
FROM `bigquery-public-data.sec_quarterly_financials.quick_summary` AS quick_summary
JOIN `bigquery-public-data.sec_quarterly_financials.sic_codes` AS sic_codes ON quick_summary.sic = sic_codes.sic_code
WHERE company_name = "BOEING CO"
AND form = "10-K"
AND LENGTH(CAST(fiscal_year AS STRING)) = 4
AND CAST(LEFT(period_end_date, 4) AS INT64) = fiscal_year
AND RIGHT(period_end_date, 4) = fiscal_year_end
AND (number_of_quarters = 0 OR number_of_quarters = 4)
ORDER BY fiscal_year
```

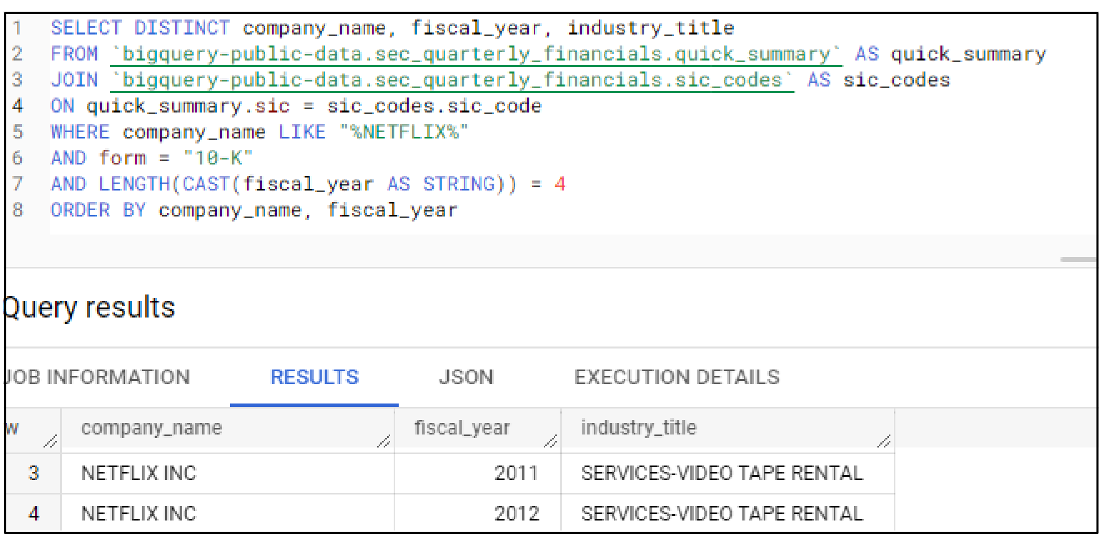
ssing location: US

ery results

INFORMATION RESULTS JSON EXECUTION DETAILS EXECUTION GRAPH PREVIEW

company_name	form	fiscal_year	fiscal_year_end	period_end_date
BOEING CO	10-K	2009	1231	20091231
BOEING CO	10-K	2010	1231	20101231

The analysis also required being able to categorize the companies by industry, but the *industry_title* field was in the *sci_codes* table within the *sec_quarterly_financials* dataset. "sic" or SIC stands for Standard Industrial Classification. To be able to combine any information in the *quick_summary* table with the *sic_codes* table, a **JOIN** function was used.



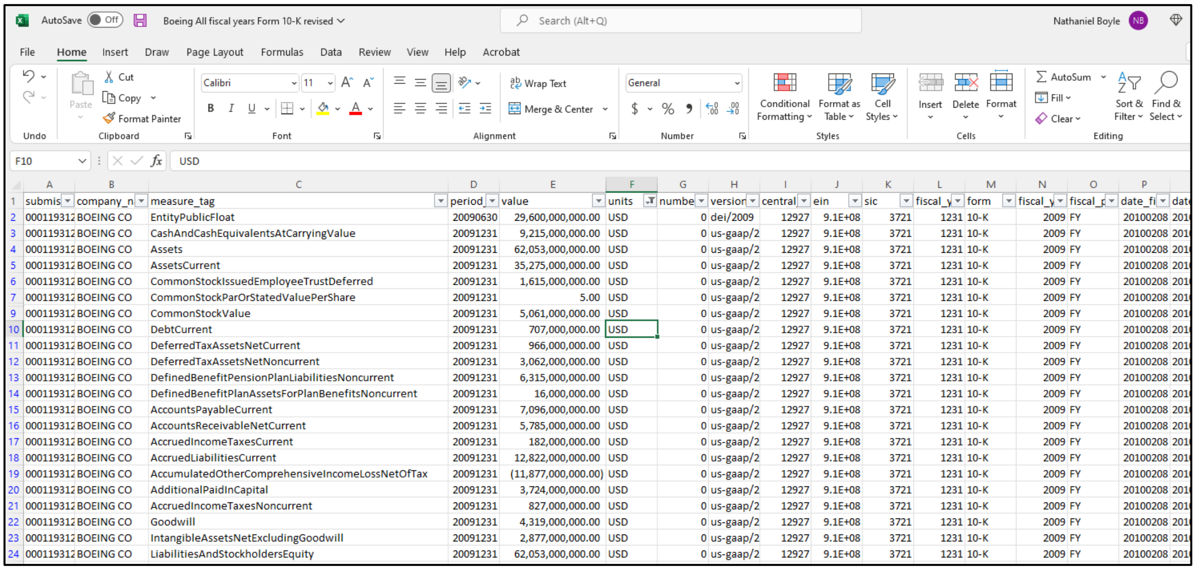
```
SELECT DISTINCT company_name, fiscal_year, industry_title
FROM `bigquery-public-data.sec_quarterly_financials.quick_summary` AS quick_summary
JOIN `bigquery-public-data.sec_quarterly_financials.sic_codes` AS sic_codes
ON quick_summary.sic = sic_codes.sic_code
WHERE company_name LIKE "%NETFLIX%"
AND form = "10-K"
AND LENGTH(CAST(fiscal_year AS STRING)) = 4
ORDER BY company_name, fiscal_year
```

Query results

JOB INFORMATION RESULTS JSON EXECUTION DETAILS

w	company_name	fiscal_year	industry_title
3	NETFLIX INC	2011	SERVICES-VIDEO TAPE RENTAL
4	NETFLIX INC	2012	SERVICES-VIDEO TAPE RENTAL

Once the tables queried had been configured as desired, they were downloaded locally in csv form and then explored in Excel to get a better, and more familiar, grasp of the data that would be worked with.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1	submis	company_n	measure_tag	period	value	units	numbe	version	central	ein	sic	fiscal_y	form	fiscal_y	fiscal_p	date_fi	date
2	000119312	BOEING CO	EntityPublicFloat	20090630	29,600,000,000.00	USD	0	dei/2009	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
3	000119312	BOEING CO	CashAndCashEquivalentsAtCarryingValue	20091231	9,215,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
4	000119312	BOEING CO	Assets	20091231	62,053,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
5	000119312	BOEING CO	AssetsCurrent	20091231	35,275,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
6	000119312	BOEING CO	CommonStockIssuedEmployeeTrustDeferred	20091231	1,615,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
7	000119312	BOEING CO	CommonStockParOrStatedValuePerShare	20091231	5.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
9	000119312	BOEING CO	CommonStockValue	20091231	5,061,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
10	000119312	BOEING CO	DebtCurrent	20091231	707,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
11	000119312	BOEING CO	DeferredTaxAssetsNetCurrent	20091231	966,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
12	000119312	BOEING CO	DeferredTaxAssetsNetNoncurrent	20091231	3,062,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
13	000119312	BOEING CO	DefinedBenefitPensionPlanLiabilitiesNoncurrent	20091231	6,315,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
14	000119312	BOEING CO	DefinedBenefitPlanAssetsForPlanBenefitsNoncurrent	20091231	16,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
15	000119312	BOEING CO	AccountsPayableCurrent	20091231	7,096,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
16	000119312	BOEING CO	AccountsReceivableNetCurrent	20091231	5,785,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
17	000119312	BOEING CO	AccruedIncomeTaxesCurrent	20091231	182,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
18	000119312	BOEING CO	AccruedLiabilitiesCurrent	20091231	12,822,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
19	000119312	BOEING CO	AccumulatedOtherComprehensiveIncomeLossNetOfTax	20091231	(11,877,000,000.00)	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
20	000119312	BOEING CO	AdditionalPaidInCapital	20091231	3,724,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
21	000119312	BOEING CO	AccruedIncomeTaxesNoncurrent	20091231	827,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
22	000119312	BOEING CO	Goodwill	20091231	4,319,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
23	000119312	BOEING CO	IntangibleAssetsNetExcludingGoodwill	20091231	2,877,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201
24	000119312	BOEING CO	LiabilitiesAndStockholdersEquity	20091231	62,053,000,000.00	USD	0	us-gaap/2	12927	9.1E+08	3721	1231	10-K	2009	FY	20100208	201

Information was also compared to the actual SEC Edgar website to make sure the data was accurate. Below are the Total Assets balances from the dataset compared to the corresponding amounts in the 2017 Boeing 10-K. (Excel on the left, 10-K on the right)

2015	94,408,000,000
2016	89,997,000,000
2017	92,333,000,000

	2017	2016	2015
Total assets	92,333	89,997	94,408

Because the raw data from the SEC Public Dataset not only would routinely use different wording for account names from one company to another, in several instances the data would use different wording for the same accounts in the same company from year to year, there were even some instances where the company name would change from one year to the next. The solution to this problem was to create keys in Excel that would catalog the account name in the dataset, verified to be the account being looked for through EDGAR, and then associate it with a standard account name. This was done for all 30 companies. Below is a screenshot of one such key. The key also contained the company names from the dataset per year, the stock exchange ticker, and the fiscal year-end. In the screenshot below the left hand column is comprised of the standard account names to be used for all accounts, to the right of that are the account names per the data set, and on the far right is a formulaic output of all accounts concatenated with quotes so they could be queried in SQL. (You may have to zoom in a bit)

		Name:	Google	
		Ticker:	GOOG	Fiscal YE: 1231
2009-2014		Company name(s):	GOOGLE INC.	
2015-2018			ALPHABET INC.	
		Measure tags:		
	A	Cash:	CashAndCashEquivalentsAtCarryingValue	"CashAndCashEquivalentsAtCarryingValue"
	B	Accounts Receivable:	AccountsReceivableNetCurrent	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent"
	C	Inventory:	InventoryNet	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet"
	D	Current Assets:	AssetsCurrent	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent"
	E	PP&E:	PropertyPlantAndEquipmentNet	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
	F	Intangible Assets:	IntangibleAssetsNetExcludingGoodwill	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
	G	Total Assets:	Assets	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
	H	Accounts Payable:	AccountsPayableCurrent	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
	I	Current Liabilities:	LiabilitiesCurrent	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
=L-K for 2009-2015	J	Total Liabilities:	Liabilities	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
	K	Stockholder's Equity:	StockholdersEquity	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
	L	Total Liabilities and Stockholder's Equity:	LiabilitiesAndStockholdersEquity	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
2009-2012	M1	Revenue:	SalesRevenueNet	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
2012-2017	M2	Revenue:	Revenues	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
2018	M3	Revenue:	RevenueFromContractWithCustomerExcludingAssessedTax	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
	N	Net Income:	NetIncomeLoss	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
2009-2014 & 2018	O	EPS:	EarningsPerShareBasic	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd
2014	P	Shares:	WeightedAverageNumberOfSharesOutstandingBasic	"CashAndCashEquivalentsAtCarryingValue", "AccountsReceivableNetCurrent", "InventoryNet", "AssetsCurrent", "PropertyPlantAnd

Once keys had been made for all 30 companies, new queries were ran for each company with the accounts chosen for analysis, this was done by industry, so three companies at a time. The below query was the generic query settled on to retrieve specific accounts from the selected companies. After the query was completed the results were saved as an industry specific table.

```
SELECT DISTINCT company_name, measure_tag AS account_name, value/1e3 AS usd_in_000s, fiscal_year, industry_title as industry
FROM `bigquery-public-data.sec_quarterly_financials.quick_summary` AS quick_summary
JOIN `bigquery-public-data.sec_quarterly_financials.sic_codes` AS sic_codes ON quick_summary.sic = sic_codes.sic_code
WHERE company_name IN ()
AND measure_tag IN ()
AND form = "10-K"
AND LENGTH(CAST(fiscal_year AS STRING)) = 4
AND CAST(LEFT(period_end_date, 4) AS INT64) = fiscal_year
AND RIGHT(period_end_date, 4) = fiscal_year_end
AND (number_of_quarters = 0 OR number_of_quarters = 4)
ORDER BY company_name, fiscal_year, account_name
```

As stated before, some companies had more than one name per year, to standardize the company names, as well as use a more user-friendly *company_name* recognizable to all, the query below was used to update company names once the respective company queries had been saved into tables in BigQuery.

```
UPDATE `shaped-repeater-314506.CapstoneProject.Banking`
SET company_name = "JP Morgan Chase"
WHERE company_name IN ("J P MORGAN CHASE & CO", "JPMORGAN CHASE & CO");

UPDATE `shaped-repeater-314506.CapstoneProject.Banking`
SET company_name = "Bank of America"
WHERE company_name IN ("BANK OF AMERICA CORP /DE/");

UPDATE `shaped-repeater-314506.CapstoneProject.Banking`
SET company_name = "Citigroup"
WHERE company_name IN ("CITIGROUP INC");
```

Partitioning of the data by industry made it easier to handle and make company and/or industry specific changes. One such change was reassigning the industry to a standard *industry* name amongst the three companies in an industry, a similar query to the one above was used to accomplish this. Another of the major changes was renaming all of the accounts to the chosen standardized names. Below is a screenshot of a few of the accounts from the query reassigning account names.

```
UPDATE `shaped-repeater-314506.CapstoneProject.Banking`
SET account_name = "Cash"
WHERE account_name IN ("CashAndCashEquivalentsAtCarryingValue");

UPDATE `shaped-repeater-314506.CapstoneProject.Banking`
SET account_name = "Accounts Receivable"
WHERE account_name IN ("AccountsReceivableNetCurrent", "AccountsReceivableGrossCurrent", "UnbilledContrac
"AccountsNotesAndLoansReceivableNetCurrent", "AccountsAndNotesReceivableNetWithNonconsolidatedAffiliates"
"UnbilledContractsReceivable");

UPDATE `shaped-repeater-314506.CapstoneProject.Banking`
SET account_name = "Inventory"
WHERE account_name IN ("InventoryNet", "InventoryFinishedGoodsNetOfReserves", "OtherInventoryNetOfReserve
"InventoryNetOfCustomerAdvancesAndProgressBillings", "InventoryNetOfAllowancesCustomerAdvancesAndProgress
;

UPDATE `shaped-repeater-314506.CapstoneProject.Banking`
SET account_name = "Current Assets"
WHERE account_name IN ("AssetsCurrent");

UPDATE `shaped-repeater-314506.CapstoneProject.Banking`
SET account_name = "PP&E"
WHERE account_name IN ("PropertyPlantAndEquipmentNet", "Land", "BuildingsAndImprovementsGross", "Property

UPDATE `shaped-repeater-314506.CapstoneProject.Banking`
SET account_name = "Intangible Assets"
WHERE account_name IN ("IntangibleAssetsNetExcludingGoodwill", "AcquiredFiniteLivedIntangibleAssetNet", "
"IndefiniteLivedIntangibleAssetsExcludingGoodwill", "IndefiniteLivedFranchiseRights", "IndefiniteLivedTra
```

The query above shows just how many account names could be used for one generic account across the 30 companies selected from the SEC dataset.

One issue that came up was that in some instances several line items in the dataset would add up to one financial line item in the 10-K. The query below was used to sum up those accounts into one (after being previously renamed) and then a table was created from that query and **UNION**'d with the original table.

```
SELECT company_name, account_name, sum(usd_in_000s) AS usd_in_000s, fiscal_year, industry
FROM `shaped-repeater-314506.CapstoneProject.all_companies_v5`
WHERE account_name = "Accounts Receivable"
GROUP BY company_name, account_name, fiscal_year, industry
ORDER BY company_name, account_name, fiscal_year
```

Another issue that came up, was that many tables were lacking key financial line items, "Total Liabilities", being one of them. But, since "Total Liabilities and Stockholder's Equity" and "Stockholder's Equity" were available, the following query used a **LAG** and appended field called *value_diff* to find the difference between the two financial line items.

```
SELECT *, usd_in_000s - LAG(usd_in_000s , 1 , 0) OVER (PARTITION BY company_name, fiscal_year ORDER BY account_name) AS value_diff
FROM `shaped-repeater-314506.CapstoneProject.all_companies_v5`
WHERE (account_name = "Stockholder's Equity" OR account_name = "Total Liabilities and Stockholder's Equity")
AND company_name = "Google"
ORDER BY company_name, fiscal_year
```

ery results

SAVE RESULTS

INFORMATION RESULTS JSON EXECUTION DETAILS EXECUTION GRAPH PREVIEW

company_name	account_name	account_type	usd_in_000s	fiscal_year	industry	value_diff
Google	Stockholder's Equity	Stockholder's Equity	36004224	2009	Information Technology	36004224
Google	Total Liabilities and Stockholde...	Total Liabilities and Stockholde...	40496778	2009	Information Technology	4492554
Google	Stockholder's Equity	Stockholder's Equity	46241000	2010	Information Technology	46241000
Google	Total Liabilities and Stockholde...	Total Liabilities and Stockholde...	57851000	2010	Information Technology	11610000

As can be seen the *usd_in_000s* (formerly the *value)* field in the *value_diff* column on the "Total Liabilities and Stockholder's Equity" row is equal to the difference between the two accounts of the same fiscal year and company. What was done next was the value in the "Total Liabilities and Stockholder's Equity" row was reassigned to the *value_diff* value, the *account_name* in that row was changed to "Total Liabilities", the *value_diff* column was dropped, the "Stockholder's Equity" rows were deleted, and the remaining rows were **UNION**'d back with the original table and a new table was saved in BigQuery. This technique was done to create other financial line items such as "Non-current Assets", "Non-current Liabilities", and "COS and Expenses". A combination of this technique and the **SUM** query up above were used to help fill in the blanks and create filler accounts such as "Other Assets" which would be a combination of accounts not large enough to have their own financial line item on the balance sheet.

Once all of the necessary changes were made and desired accounts were included in the final dataset, the next, and arguably most interesting, part of the project could begin.

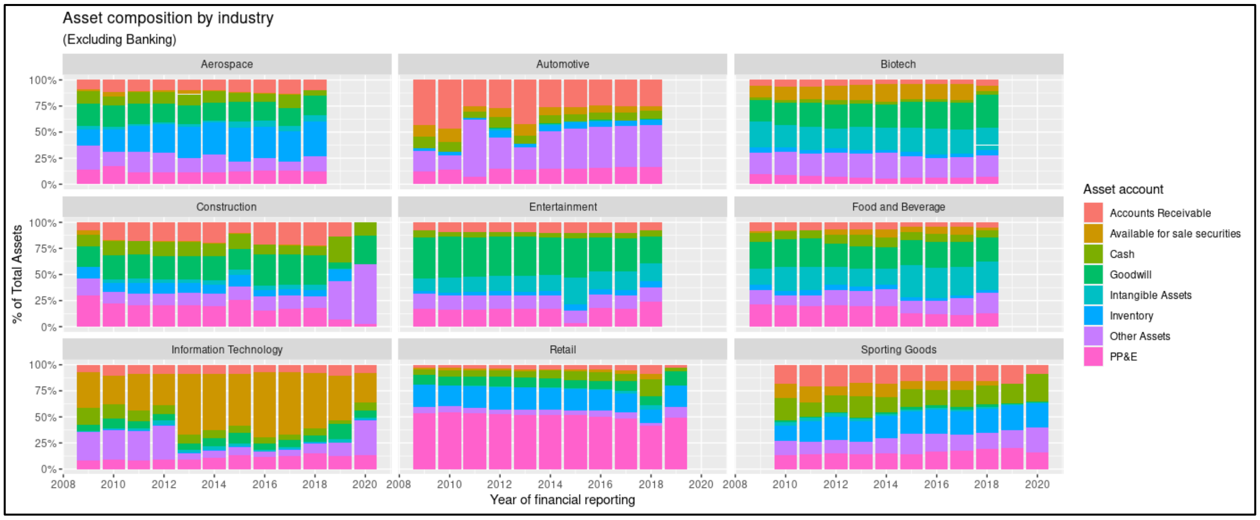
III. Data Analysis with R Studio

Data Analysis of the now formatted data was performed using "R" a programming language for statistical computing and graphics and was done in the software environment for the language known as "R Studio". For each section below there will be commentary on the main chart/plot, and depending on how unorthodox, interesting, or important the industry specific findings are, there will be additional detailed commentary.

Asset Composition (for R code please see Appendix **A** at the end of this document)

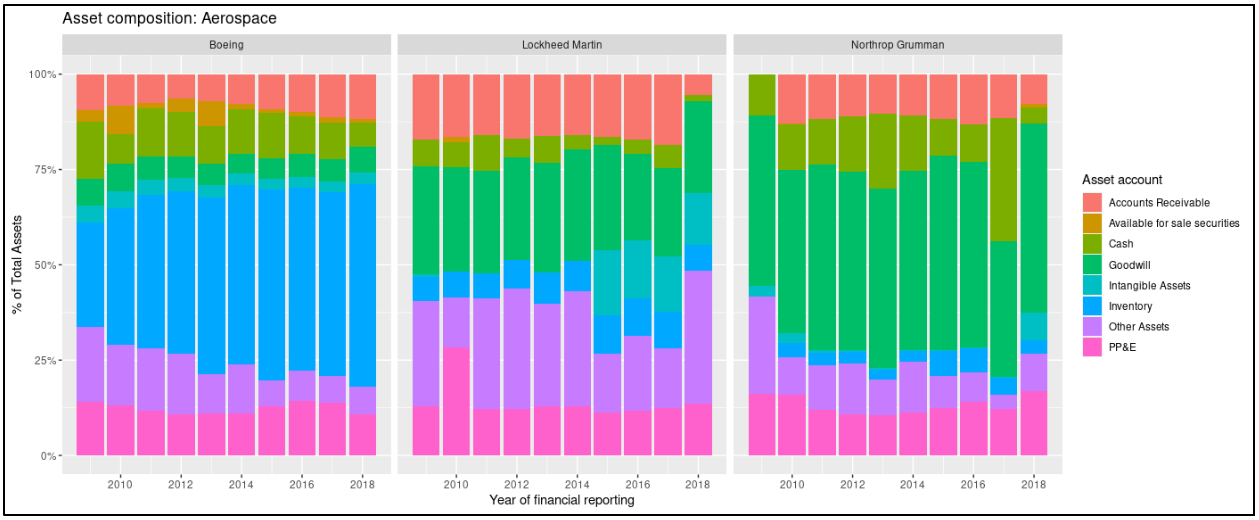
When planning out this project the original idea was to mainly use line and dot plots, and to show growth of some kind over time, but given the nature of the data and how accounting systems and financial statements are structured, as the data was dug into more it became clear that a percent stacked bar chart would be better for visualizing the relationships between many if not most accounts. Below we see a percent stacked bar chart of asset composition by industry.

Assets by industry



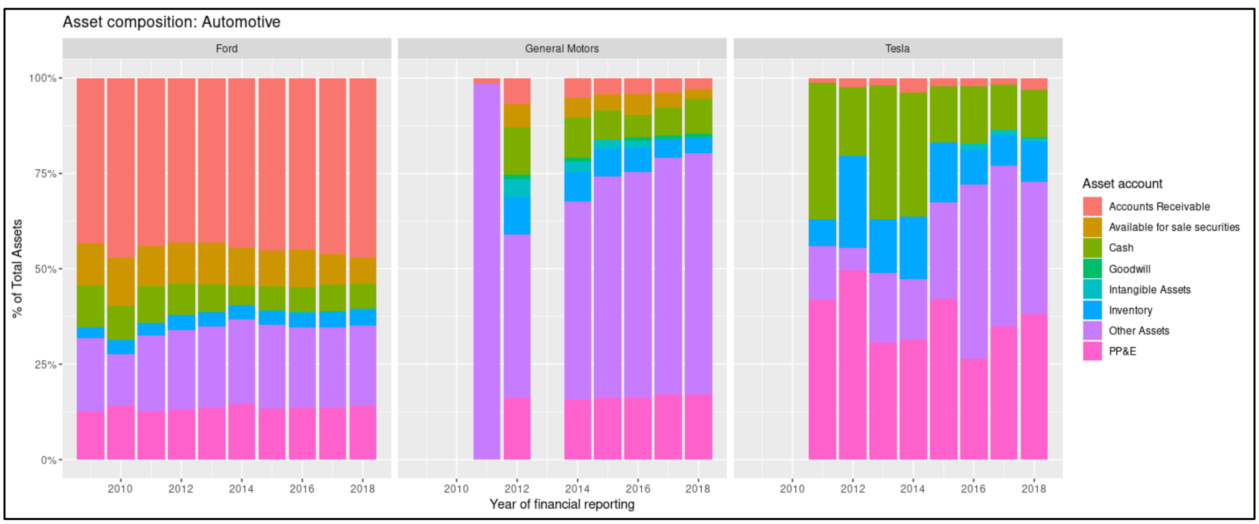
What was initially surprising was the percent of "Available for sale securities" that made up the Total Assets of information technology companies. Although, when it is taken into consideration the large amount of cash these companies have on hand it would reason that they would store it in an appreciable asset. Something else that was interesting, but not quite as surprising as the Available for sale securities finding, was just how large of a percent "PP&E" (Property, Plant, and Equipment) was for the Total Assets of the retail industry, but when taken into account how much real estate, for store fronts or warehouses those companies would us to sell and store inventory, and the equipment used in those warehouses as well as the trucks needed to move inventory, it begins to make a bit more sense why PP&E was so large, still interesting though how much of a greater percent it is than "Inventory". Speaking of Inventory, another bit of a surprise is the percent Aerospace industry assets that is made up of Inventory, in the industry specific break downs in the subsequent charts a more detailed explanation of why is given. Finally, the last item that will be commented on here before the industry specific break down is how large of a percent "Accounts Receivable" makes up of Automotive assets, although when considering how cars are financed for buyers, it would seem logical.

Aerospace Assets



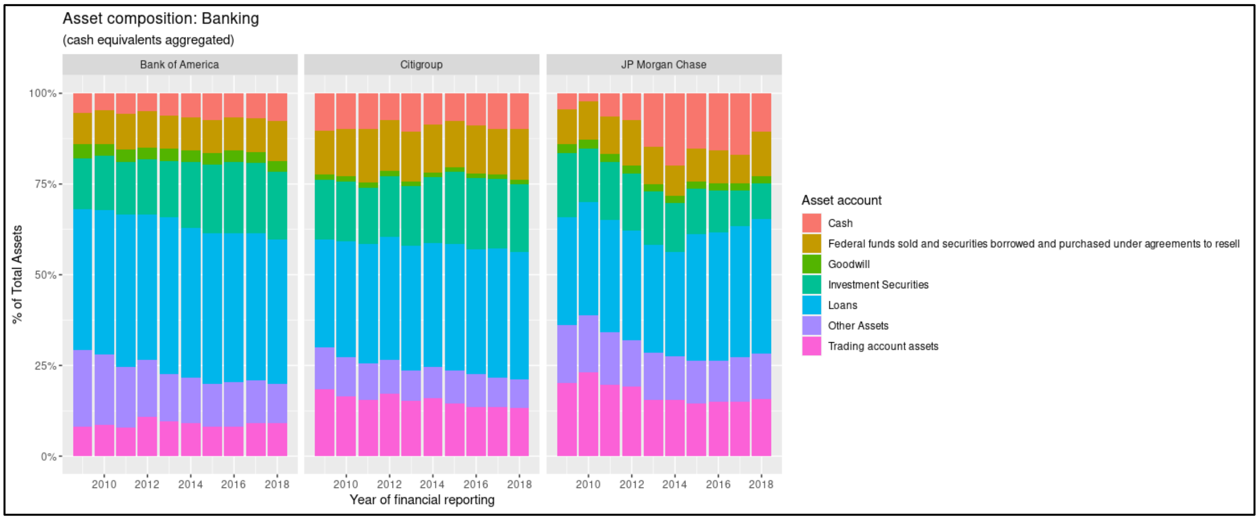
From the graph above it appears that Boeing was the main contributor to the aerospace industry having such a high Inventory percentage. This is most likely due to all of the commercial aircraft they manufacture, as many airliners and cargo planes are of Boeing make.

Automotive Assets



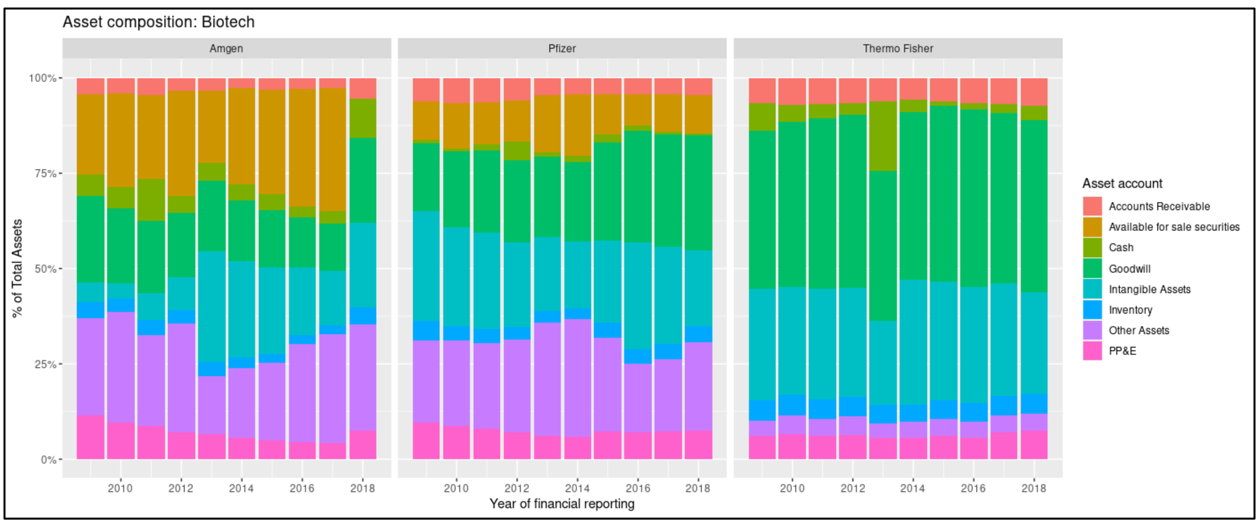
As stated earlier in Part II., the data that came from the SEC dataset was extremely raw, and sometimes unreliable, here we see missing years and accounts for General Motors, and later on it will be shown that the Ford data was lacking the “Current Assets” and “Current Liabilities” financial statement line items for some years. As far as Ford is concerned here though, it is likely the most accurate asset breakout of a mainstream automotive company in our dataset. Ford's asset composition has a decent percentage of PP&E and the plurality of the assets are Accounts Receivable which would make sense given it is an established brand with dealerships, or cars at dealerships, virtually everywhere in the United States, thus there would be a lot of Ford cars in the process of being paid for. For Tesla, the large PP&E percentage makes sense, given it is a new company and has invested heavily in new infrastructure such as manufacturing plants, the size of the Cash balance compared to other assets is a bit of a surprise, but is most likely due to the various government loans and subsidies the company received after going public in 2010.

Banking Assets



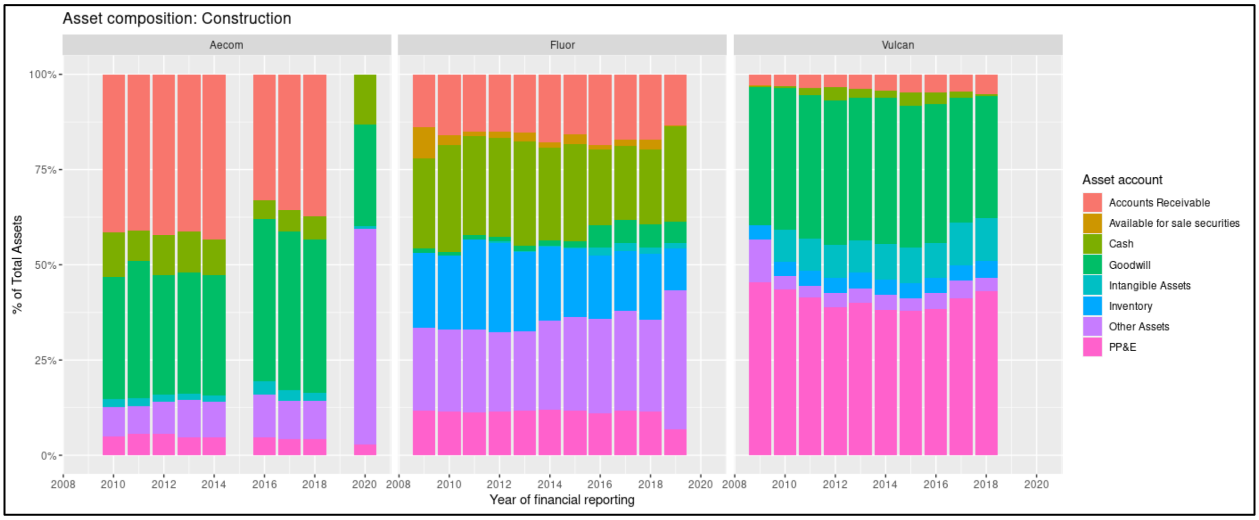
Nothing overly surprising about the Banking asset composition, at least as far as banking goes, the asset make up is certainly quite different from the industries. Otherwise, the percentages seem to be fairly uniform across the three banks. In another section further down the document, the role some of these assets play in certain banking ratios will be demonstrated.

Biotech Assets

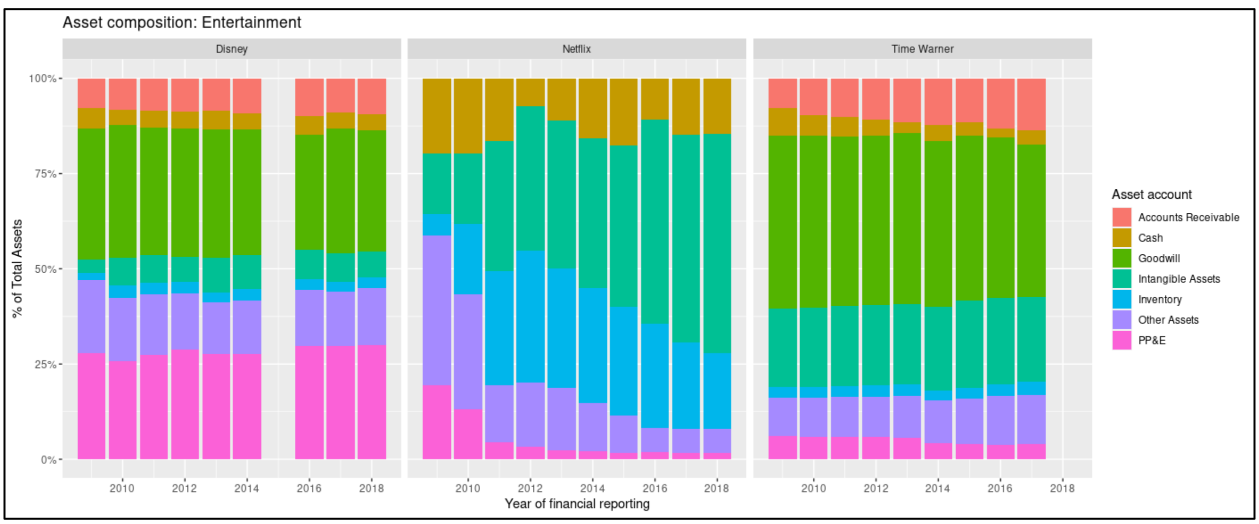


Biotech, at least Amgen and Pfizer, seems to have a decent amount of investment in securities, given the unreliable nature of products making it passed FDA trials and making it to market, it may be wise to hedge and invest the cash a company has on hand. One asset group in a large percentage here that is not surprising to see is "Intangible Assets". Here these Intangible Assets are most likely patents on certain pharmaceutical products and/or medical devices, this would be the primary source of revenue for a biotechnology company (before the patent expires) and thus one would expect a large portion of the value of their assets to be Intangible Assets.

Construction Assets

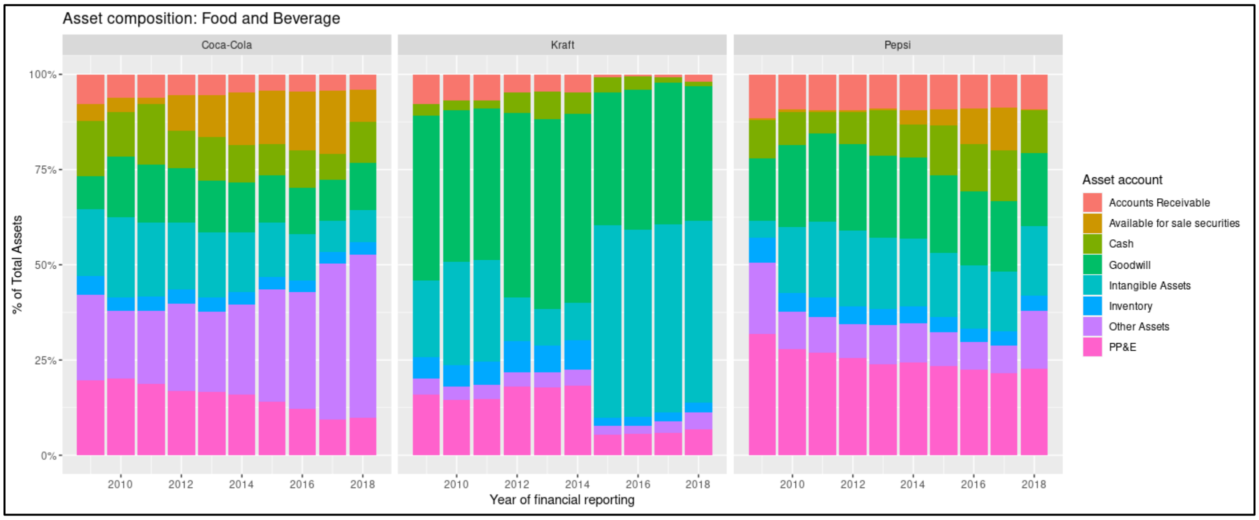


Aecom seems to be more Accounts Receivable, Vulcan more PP&E, with Fluor landing somewhere in the middle. Given that Aecom is more of a construction consulting firm, a large receivables balance compared to fixed assets makes sense. Vulcan is more downstream than the other two companies, and actually produces the materials construction companies use, so it would reason that a large PP&E balance would be seen here. Fluor provides actual engineering and construction as well as project management, so a relative blend of assets here is reasonable.



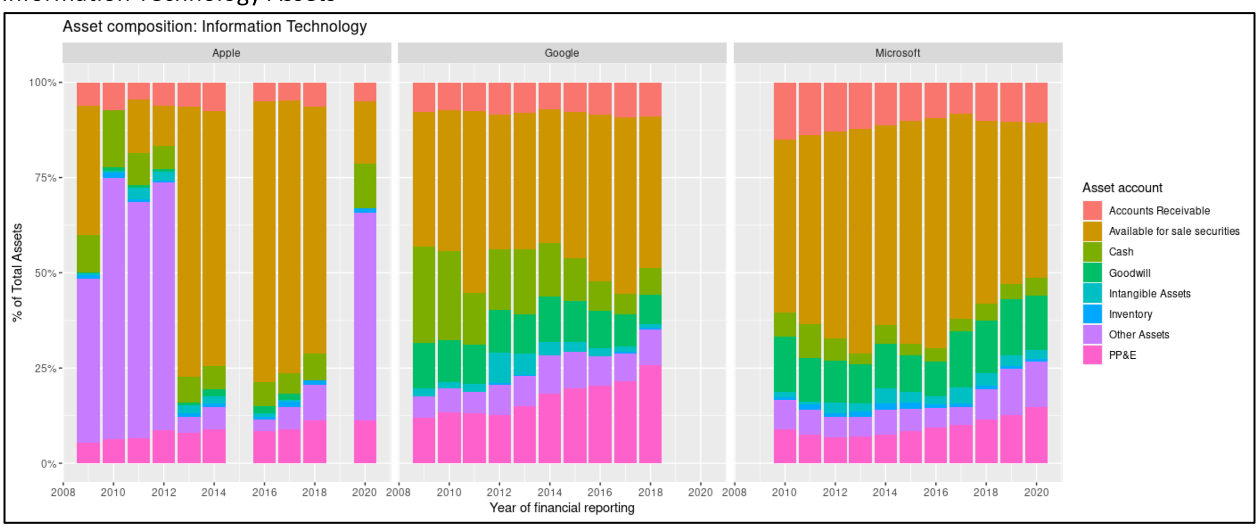
Disney's large PP&E balance is most likely due to the parks it owns and operates. Time Warner's large Intangible Assets balance is likely the result of all of the acquisitions it has been a part of over the years. For Netflix, it is a bit surprising to see virtually no Accounts Receivable balance, but when the subscription model is taken into consideration it begins to become clear as to why, basically when ever a subscriber pays their monthly due it is essentially a direct deposit into Netflix's account, so at most a customer could only have one month's of a receivable balance with the company before either the balance was paid or the subscription was put on hold. The large Intangible Asset balance is likely the computer software unique to and developed by Netflix for their streaming service, and the large Inventory balance is actually their current content assets (film & TV library) at the time of filing.

Food and Beverage Assets



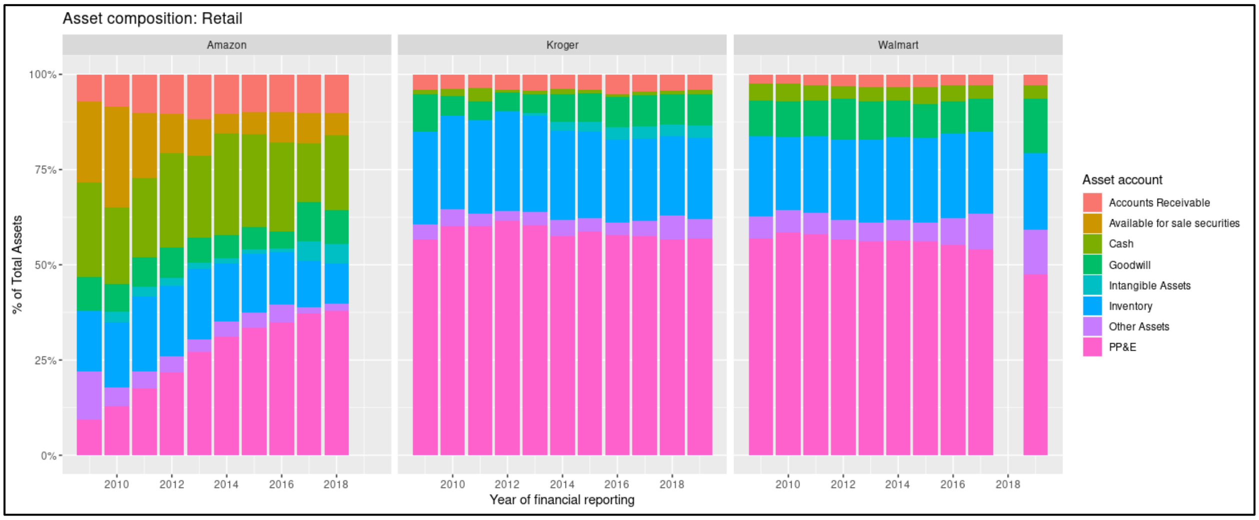
Coca-Cola and Pepsi appear to be fairly uniform in relation to each other, the larger PP&E balance from Pepsi however, is likely due to the restaurant chains owned by Pepsi, such as KFC or Taco Bell which would require real estate, buildings, and equipment to operate. The large spike in Intangible Assets seen in Kraft is due to the Kraft-Heinz merger that occurred in 2015 and the decrease seen in 2012 is from Kraft Foods, Inc. spinning off its North American grocery business into a new company called Kraft Foods Group, Inc., very creative name.

Information Technology Assets



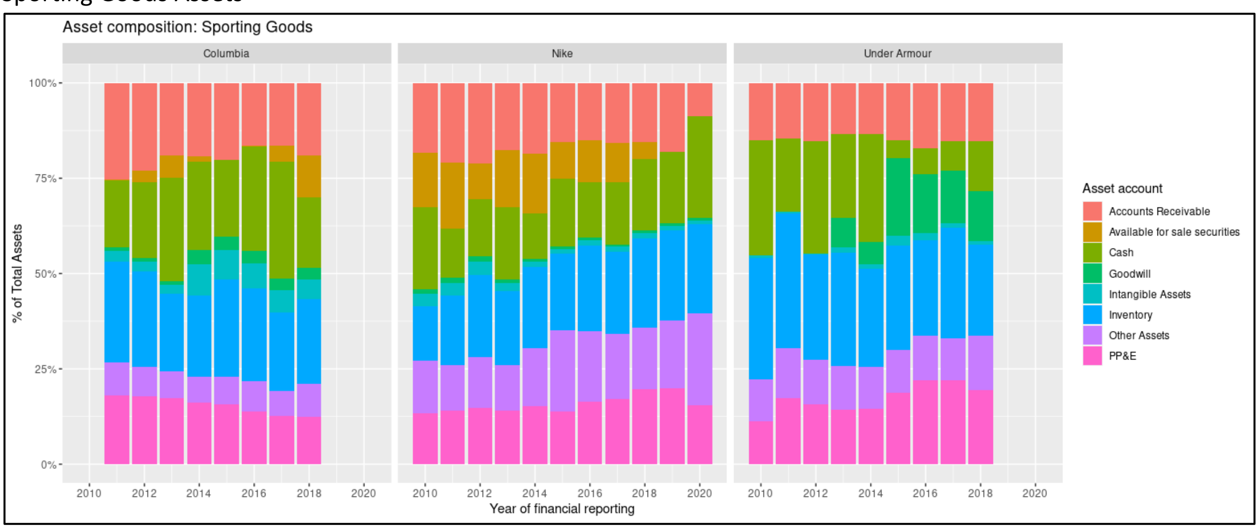
As covered in the preliminary analysis at the beginning of this section for Information Technology companies there is a disproportionate amount of the Total Assets balance being comprised of Available for sale securities, but again, given the extraordinary amount of cash inflow these companies receive it is probably wise to invest that money into assets that appreciate at a rate higher than whatever interest rate their banks would be offering on deposits.

Retail Assets



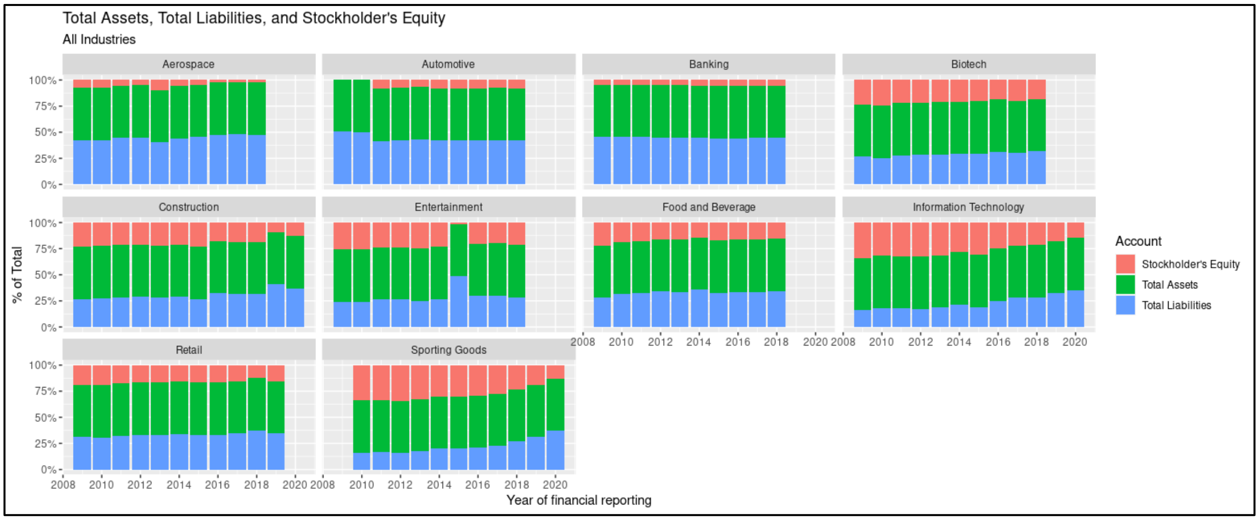
Given the real estate and equipment requirements to move and store large inventories, a large PP&E balance is not surprising for the three retail companies chosen for analysis, neither is a large Inventory balance. Amazon is a bit of a hybrid between an information technology company and a retailer, and as is seen here Amazon's asset makeup slowly starts to relatively become less comprised of securities (like other IT companies) and more so of PP&E (like the other retailers), this is most likely due to not only the addition of warehouses, but also server farms for AWS (Amazon Web Services) as well as Amazon creating their own fleet of shipping trucks.

Sporting Goods Assets



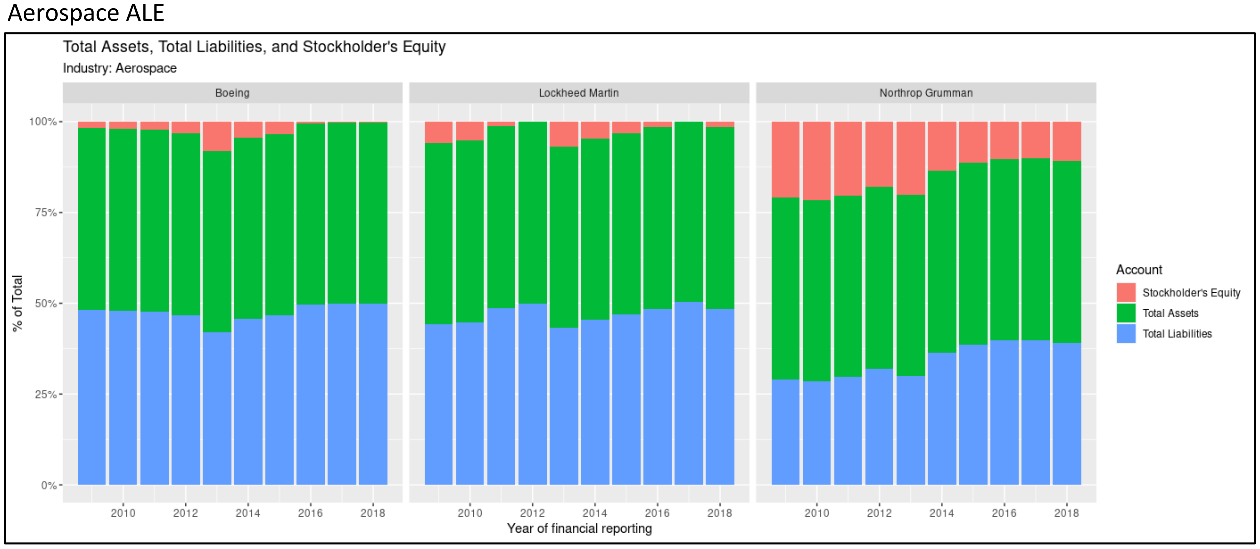
Last but not least, sporting goods. From the chart above it can be seen that there is fair amount of uniformity between the three companies, with decent proportions of Accounts Receivable, Cash, Inventory, and PP&E, all accounts that would be expected to play a large role in the operations of a sporting goods company.

Asset, Liability, and Equity (for R code please see Appendix **E** at the end of the document)



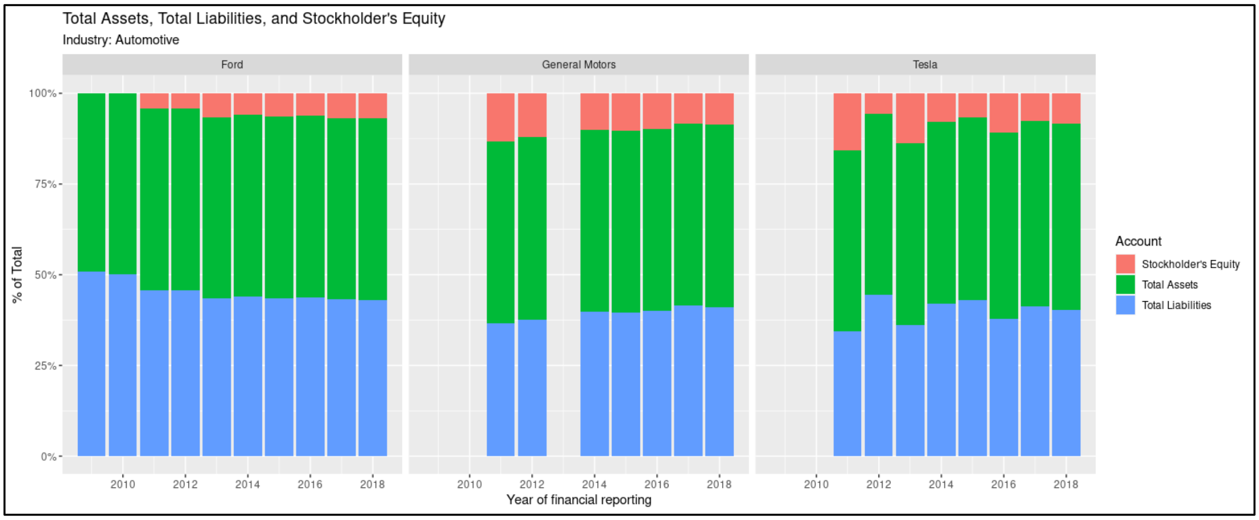
The classic A = L + E accounting equation brought to life. There aren't too many interesting insights to be gathered from these charts, but it is a bit fascinating that rather than just seeing a number to represent an Assets to Equity, Liabilities to Equity, or Liabilities to Assets ratio, instead what is shown is a crisp visualization of how Assets, Liabilities, and Equity all compare to each other. One interesting insight that was gathered from this analysis, was that some companies actually had negative Equity, and that they actually had more Liabilities than Assets, the structure of these graphs cannot fully demonstrate negative amounts, but two years of filings for Ford and one for Lockheed Martin had negative Equity, it should be pretty easy to spot which years in the graphs below.

Aerospace ALE

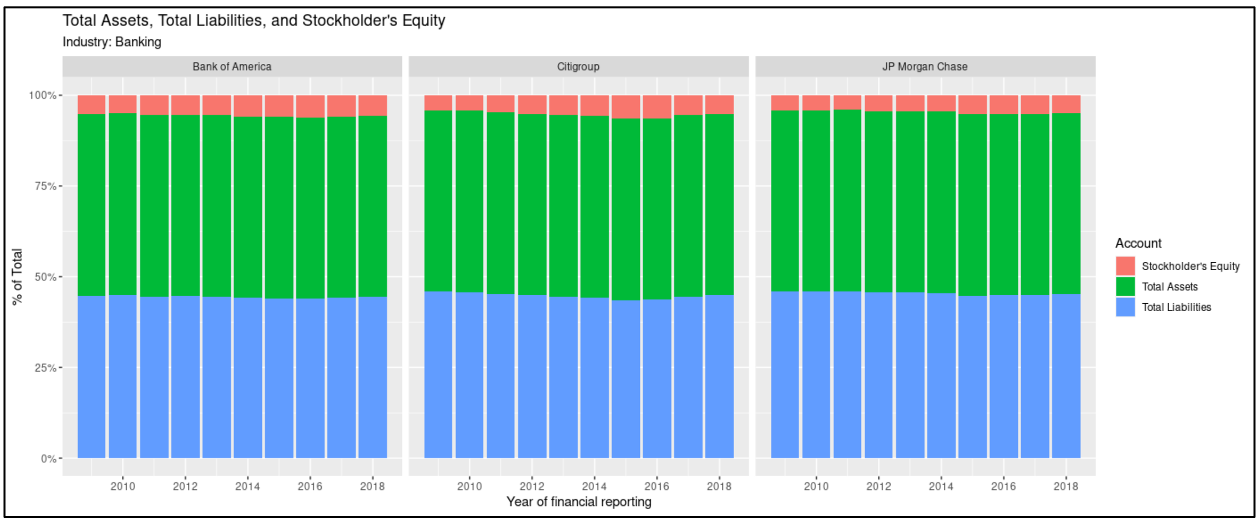


For the industry specific chart above and charts below, no commentary will be provided as, for the most part, the results are fairly straightforward, but the graphs are still a neat visualization to see, thus their inclusion.

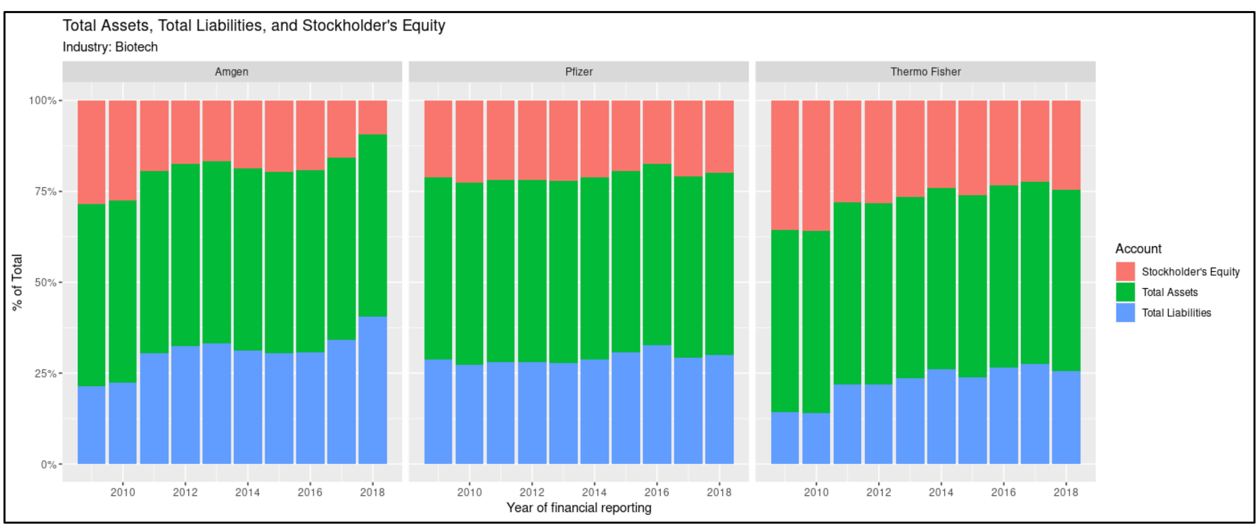
Automotive ALE



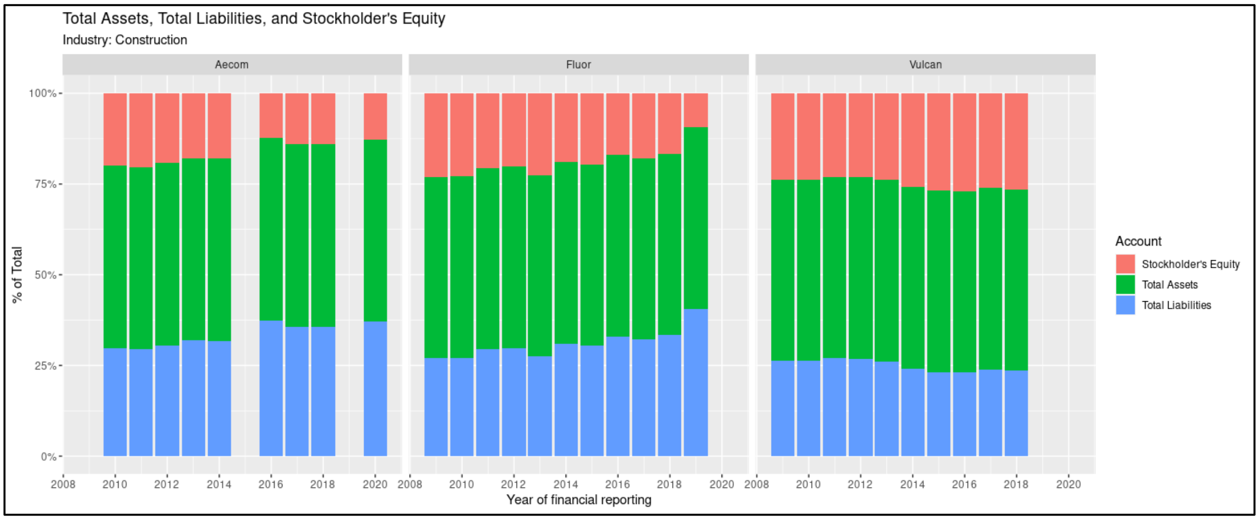
Banking ALE



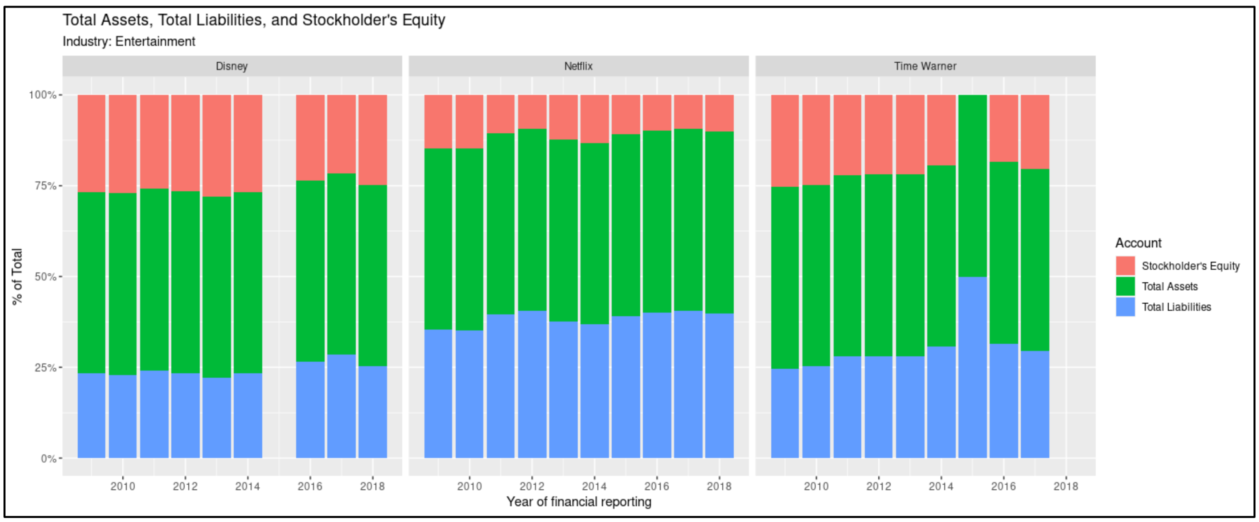
Biotech ALE



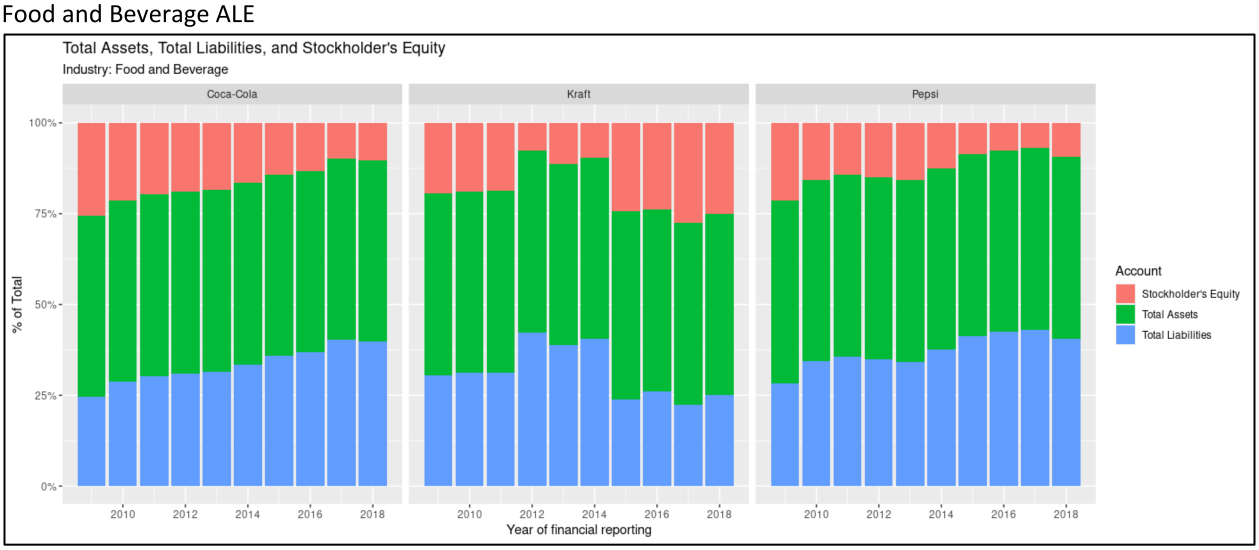
Construction ALE



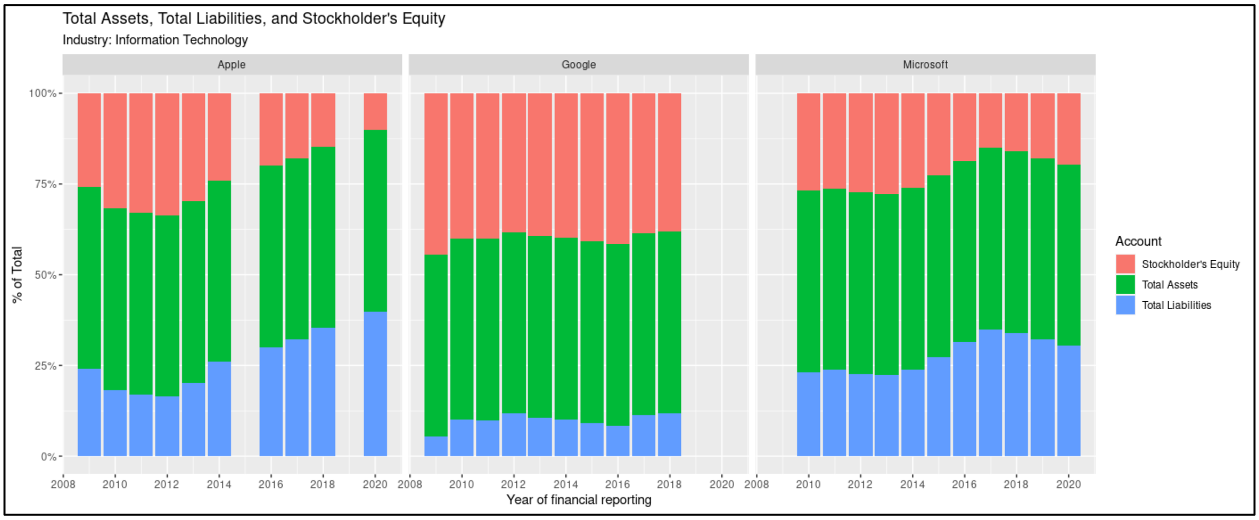
Entertainment ALE



Food and Beverage ALE

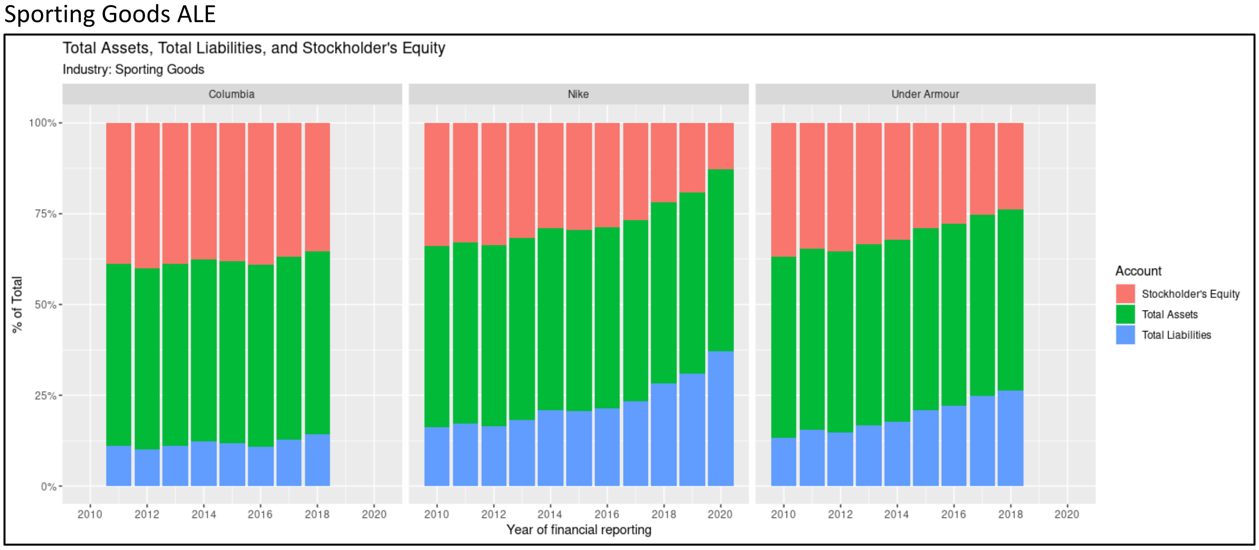


Information Technology ALE

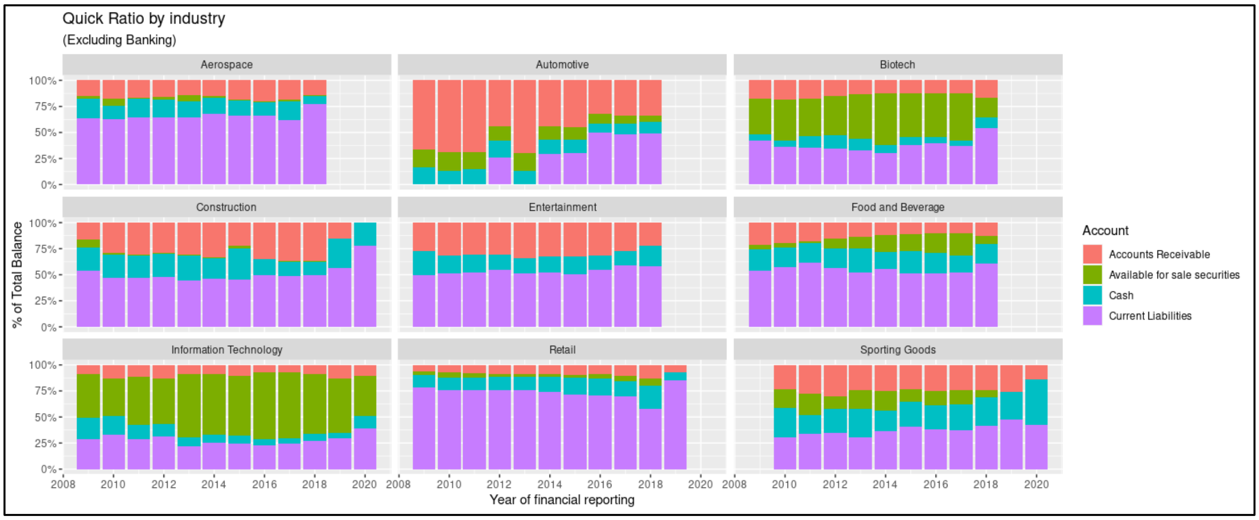


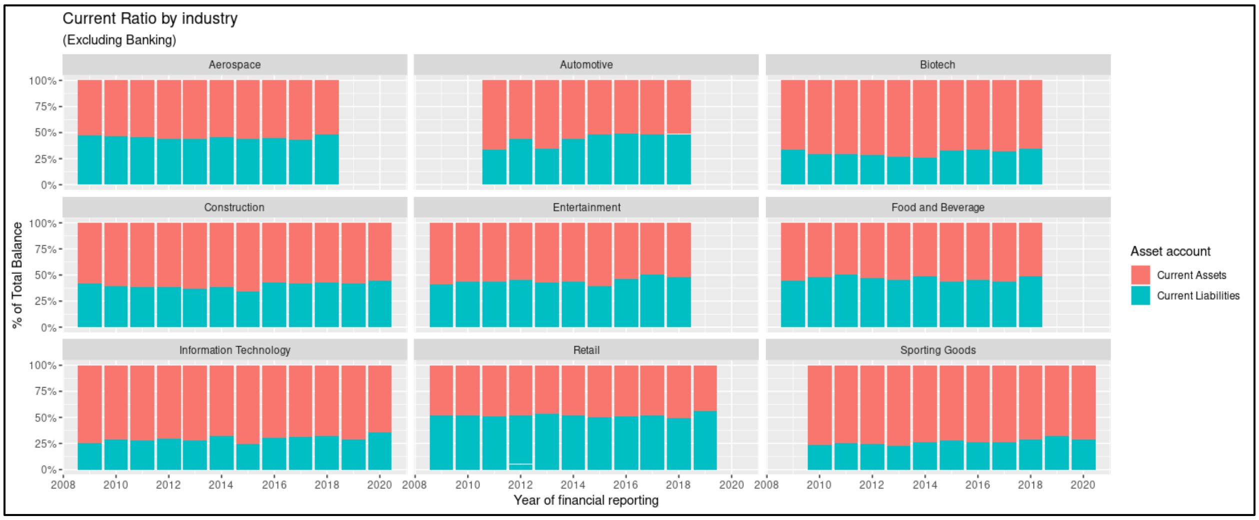
Retail ALE

Sporting Goods ALE



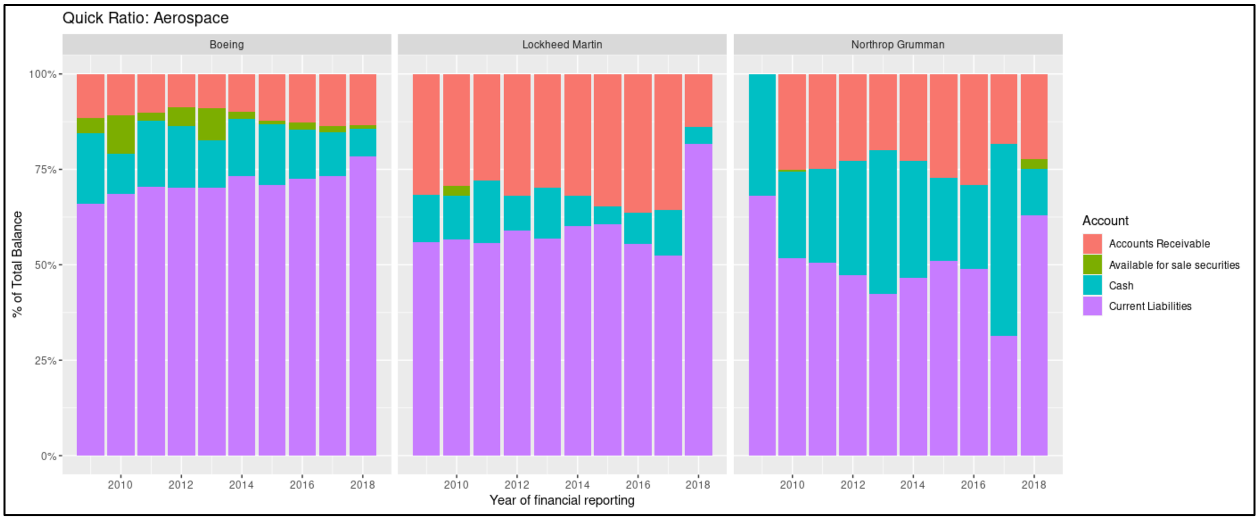
Liquidity Ratios (for R code please see Appendix **D** at the end of the document)

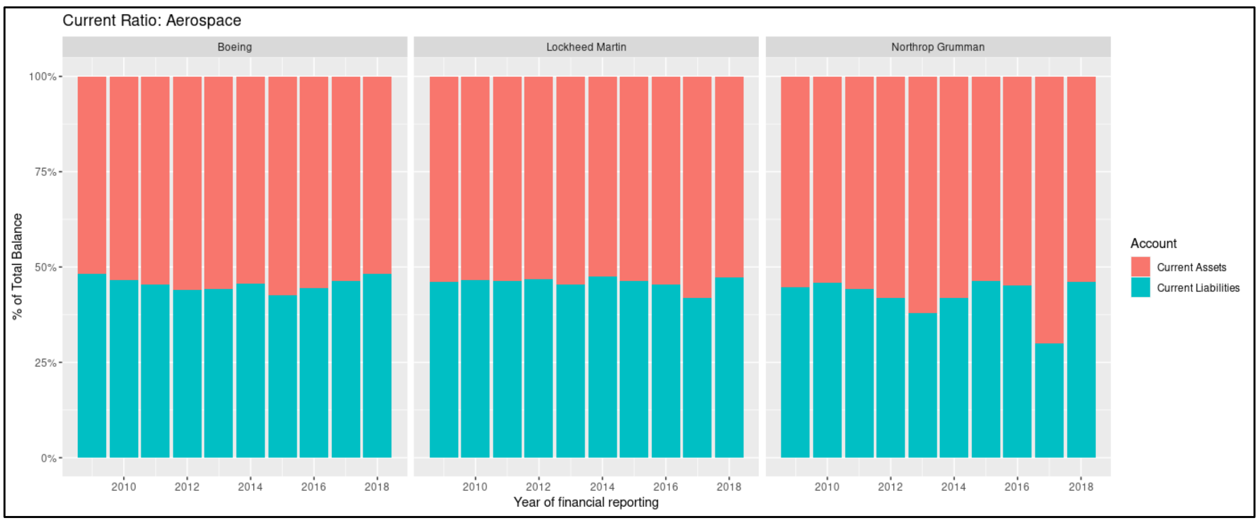




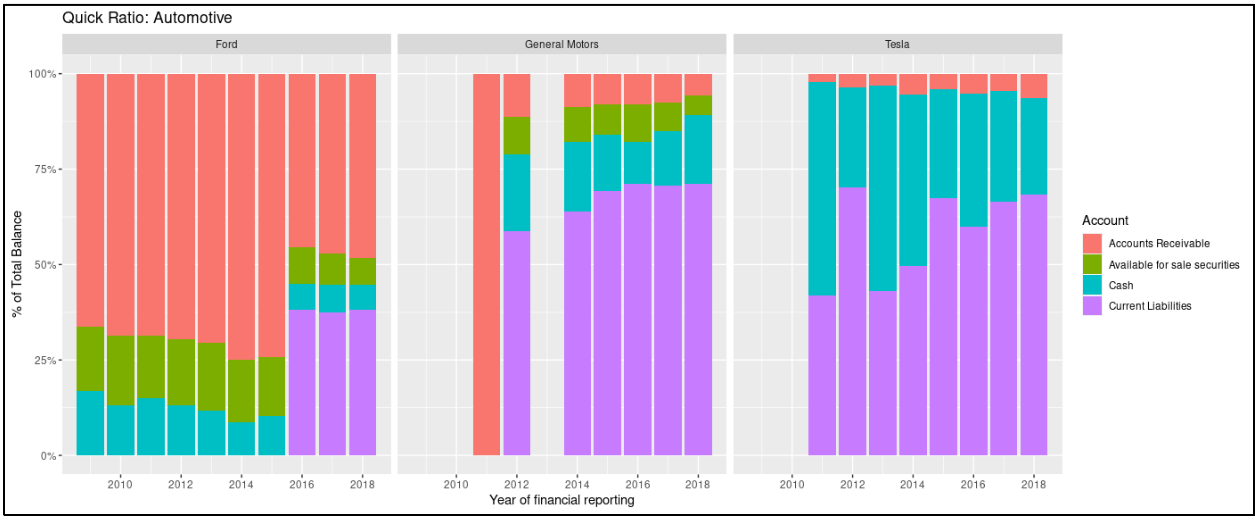
The Quick Ratio is usually defined as Current Assets, less Inventory and Prepaid Assets, divided by Current Liabilities, while the Current Ratio is just Current Assets divided by Current Liabilities. Both of these ratios are used as measures of a company's liquidity (its ability to pay off short-term obligations or liabilities). One interesting insight to take from the charts up above and the industry specific charts down below is how much of a role inventory can play in a company's liquidity depending on the type of industry it is in, this is why the Quick Ratio and Current ratio charts were displayed next to each other instead of in separate series.

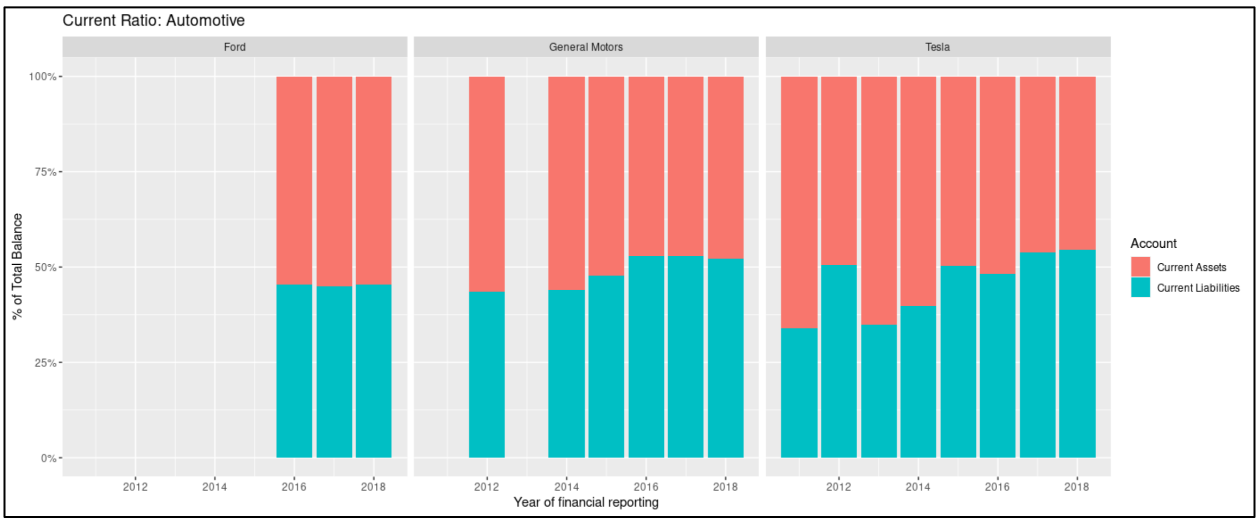
Aerospace liquidity ratios:



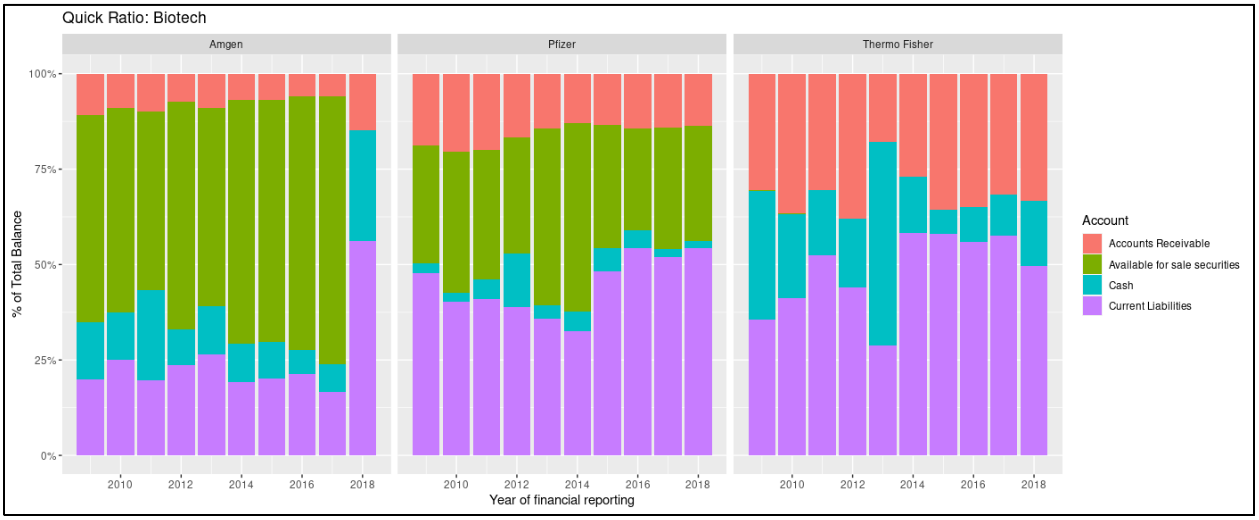


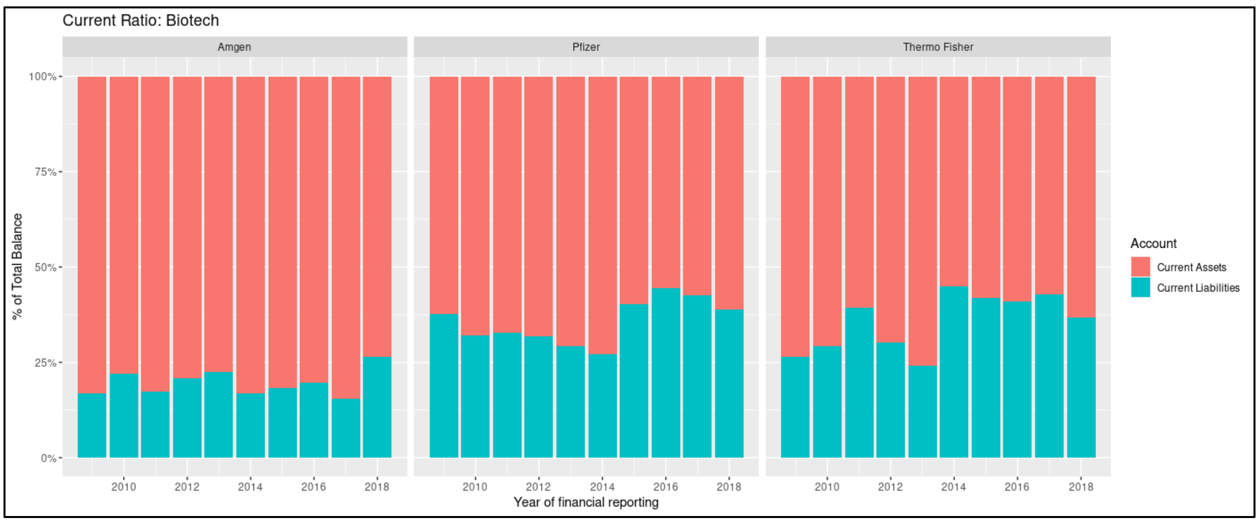
Automotive liquidity ratios:



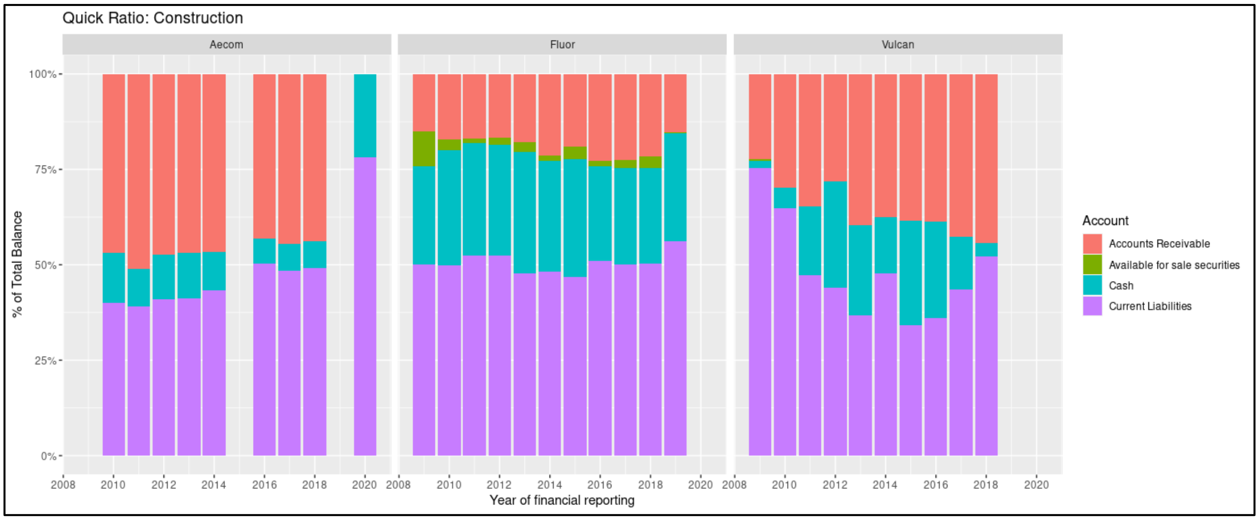


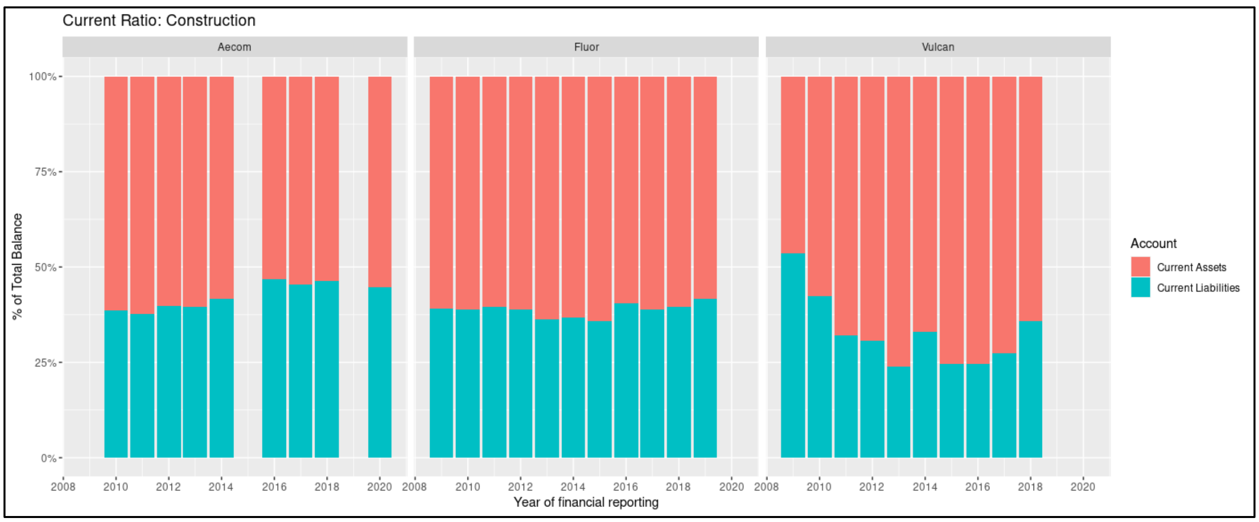
Biotech liquidity ratios:



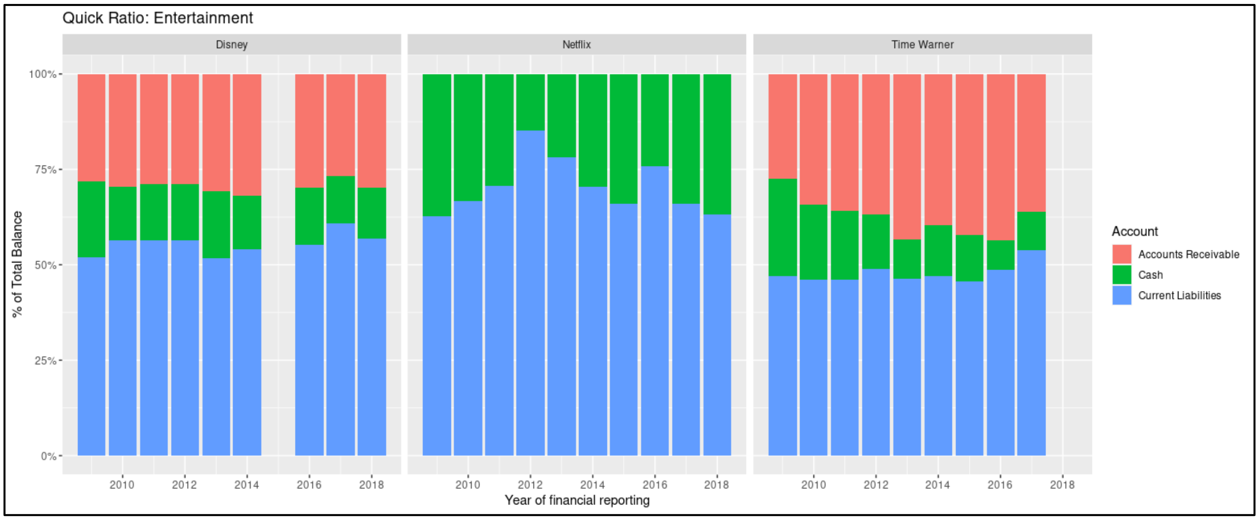


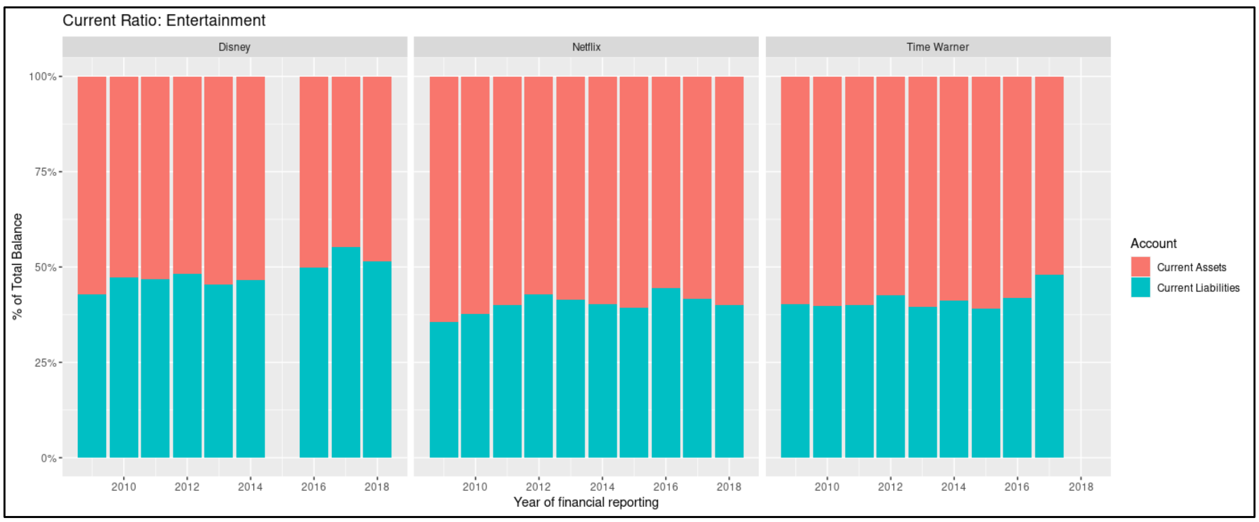
Construction liquidity ratios:



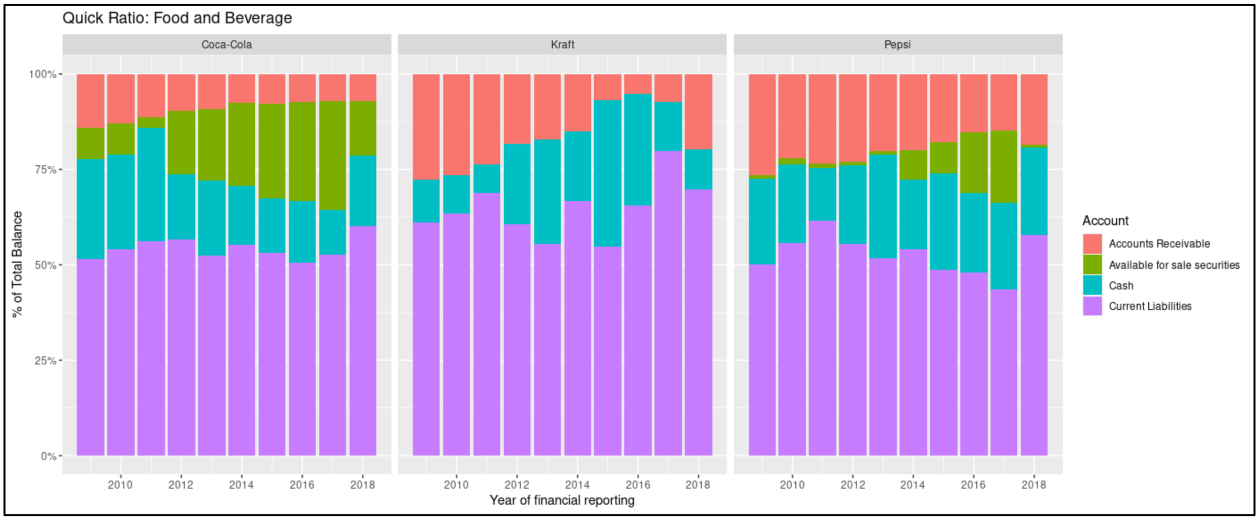


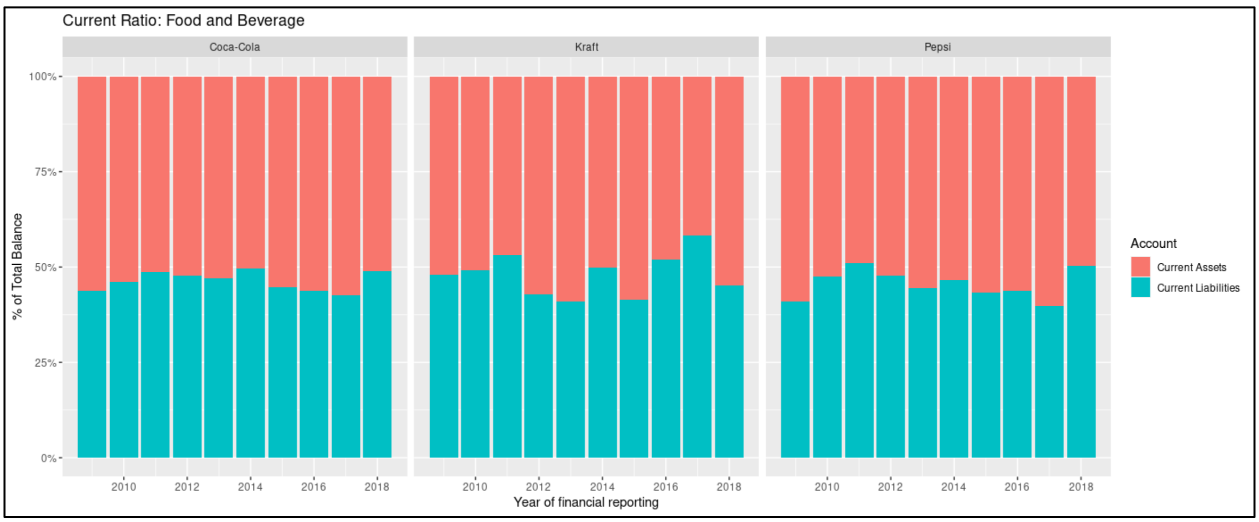
Entertainment liquidity ratios:



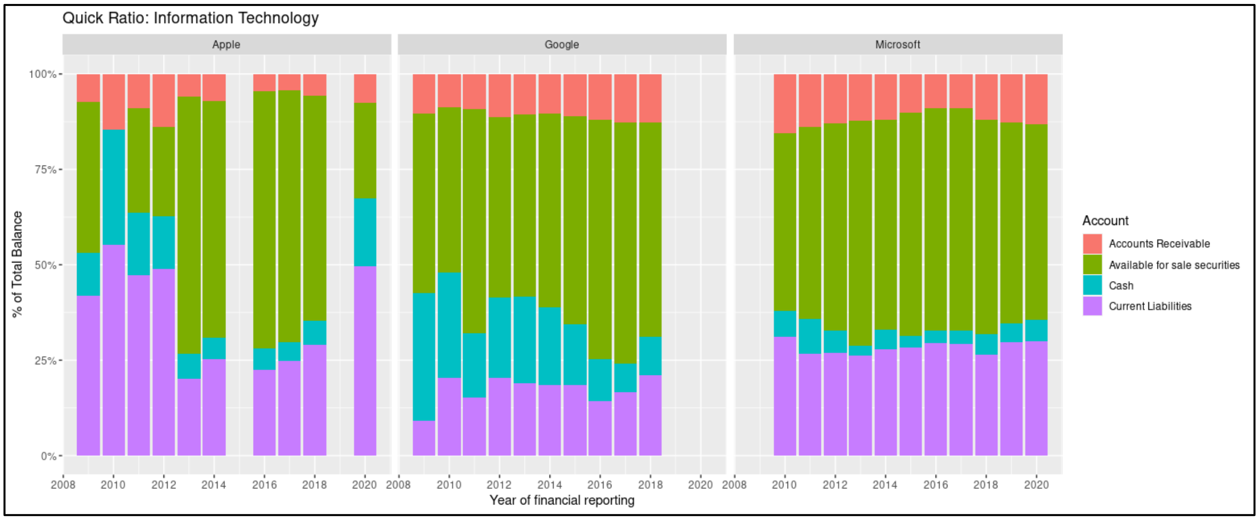


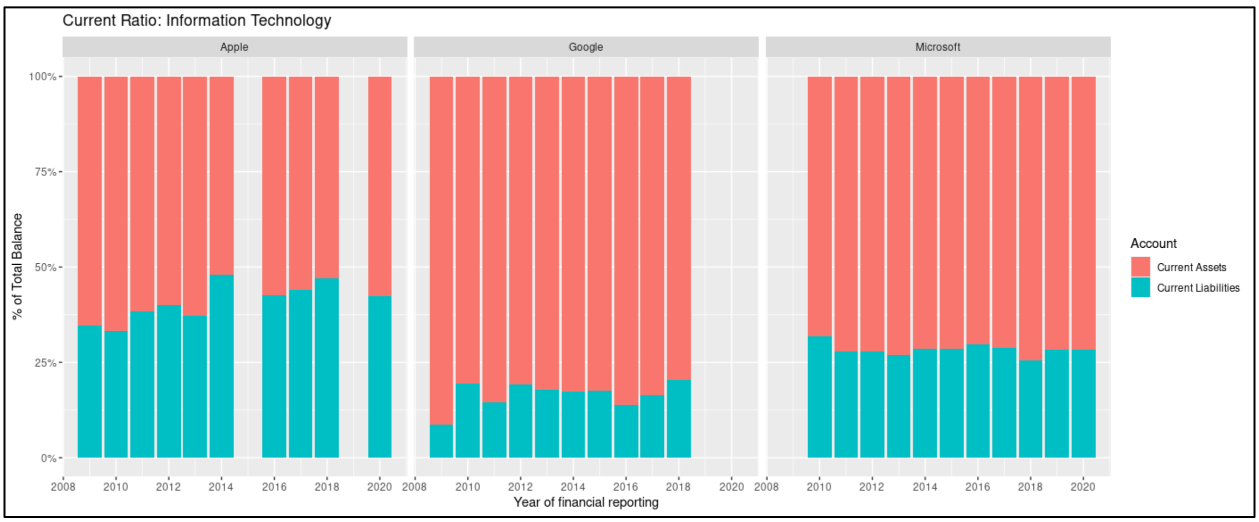
Food and Beverage liquidity ratios:



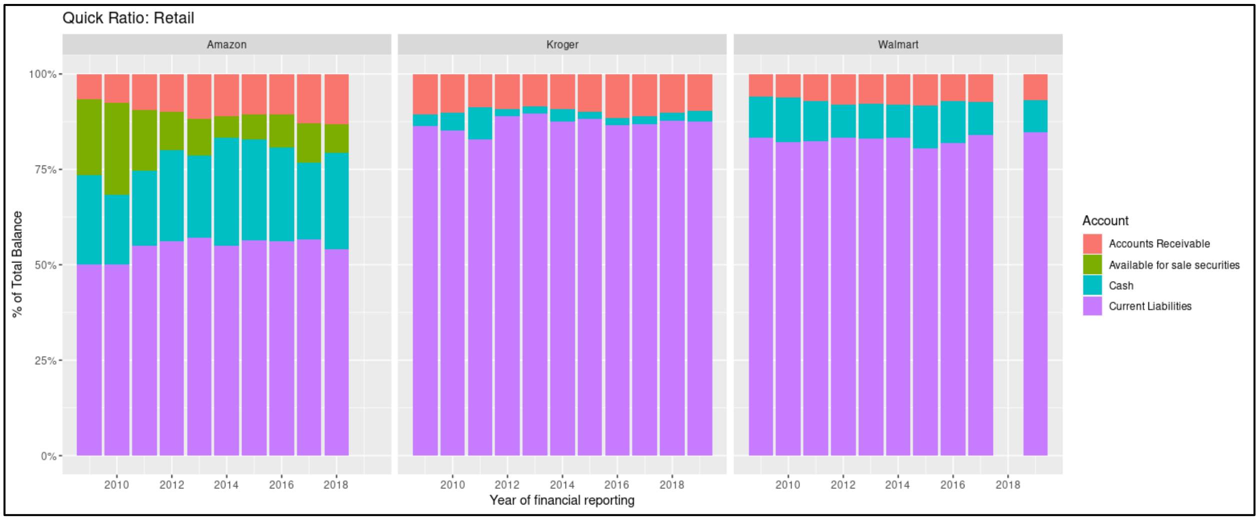


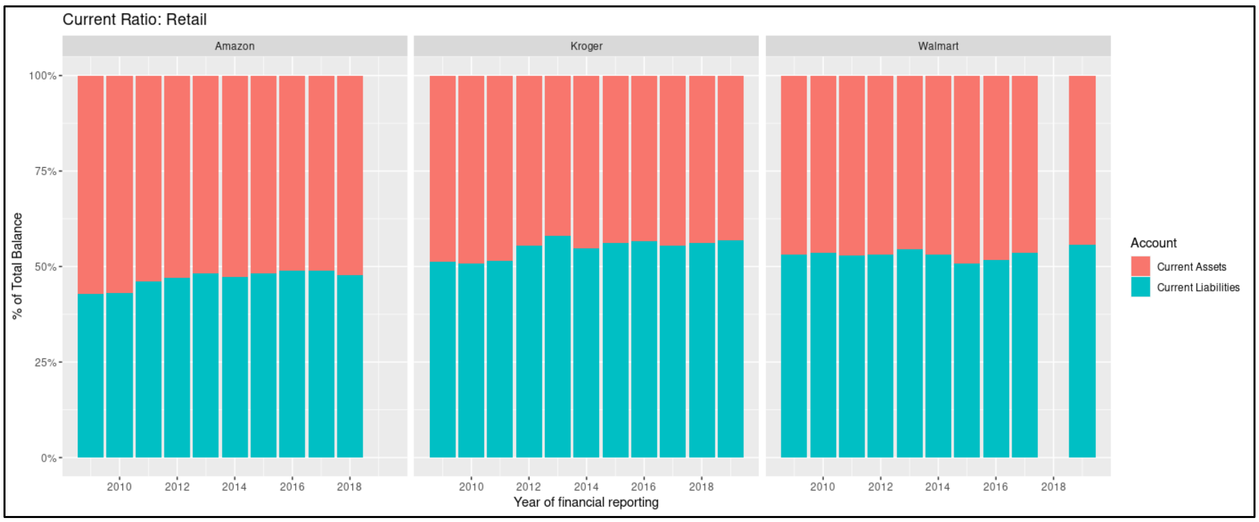
Information Technology liquidity ratios:



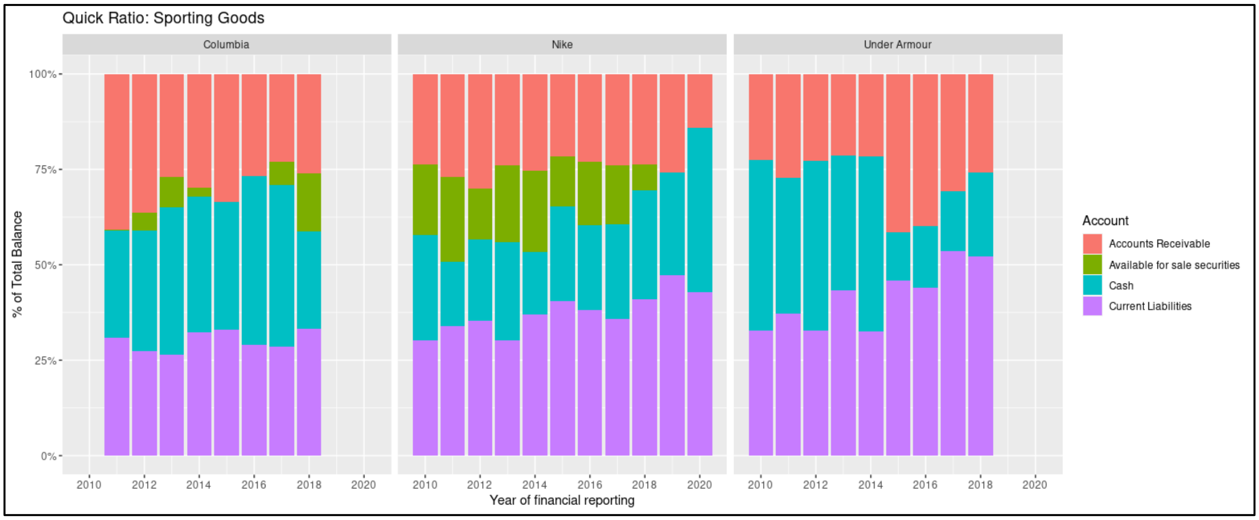


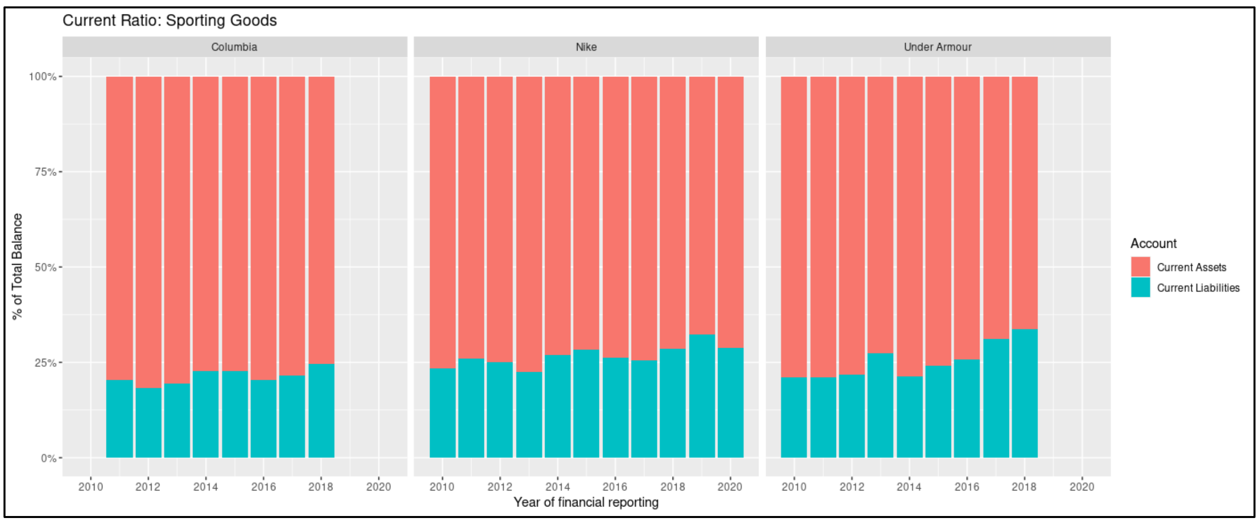
Retail liquidity ratios:



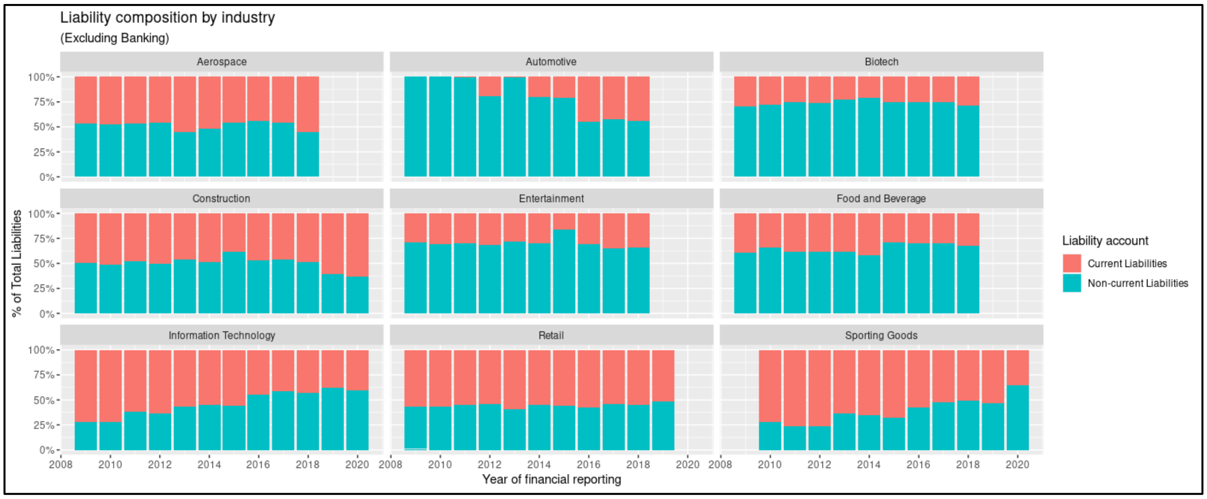


Sporting Goods liquidity ratios:



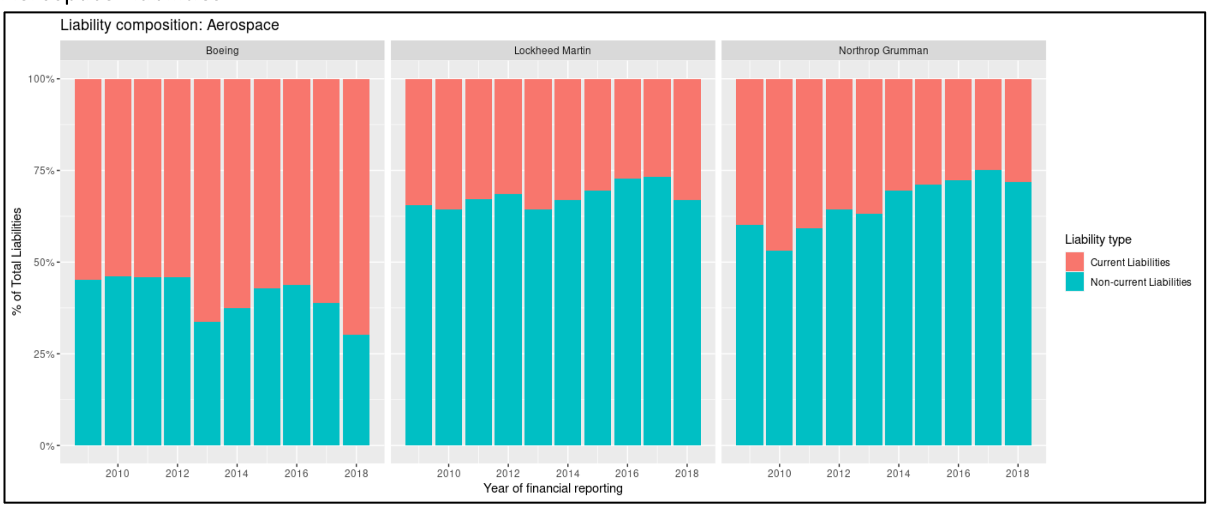


Current and Non-current Liabilities <u>(for R code please see Appendix **F** at the end of the document)</u>

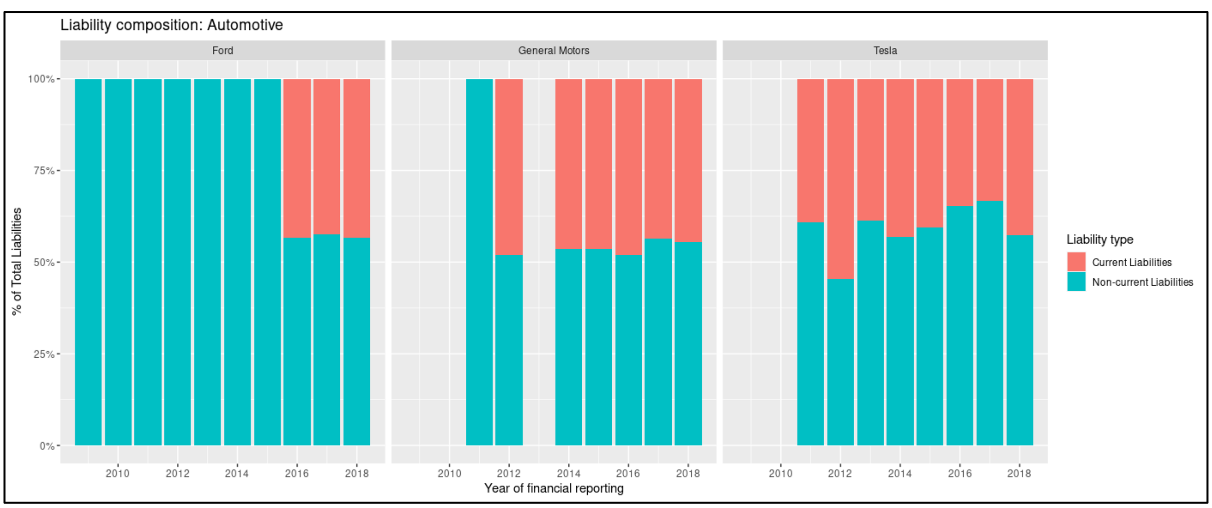


There isn't a whole lot of interesting insights to be gained from this chart or the industry specific charts below, they were included just to show how debt is leveraged across the various industries and companies and which is more likely to take on short-term vs. long-term obligations.

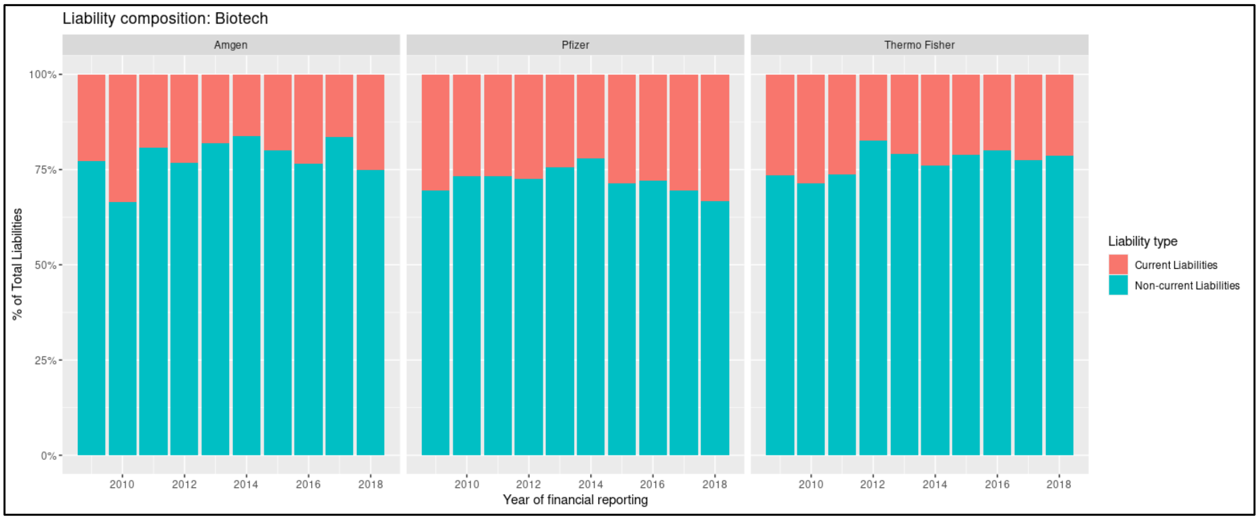
Aerospace Liabilities



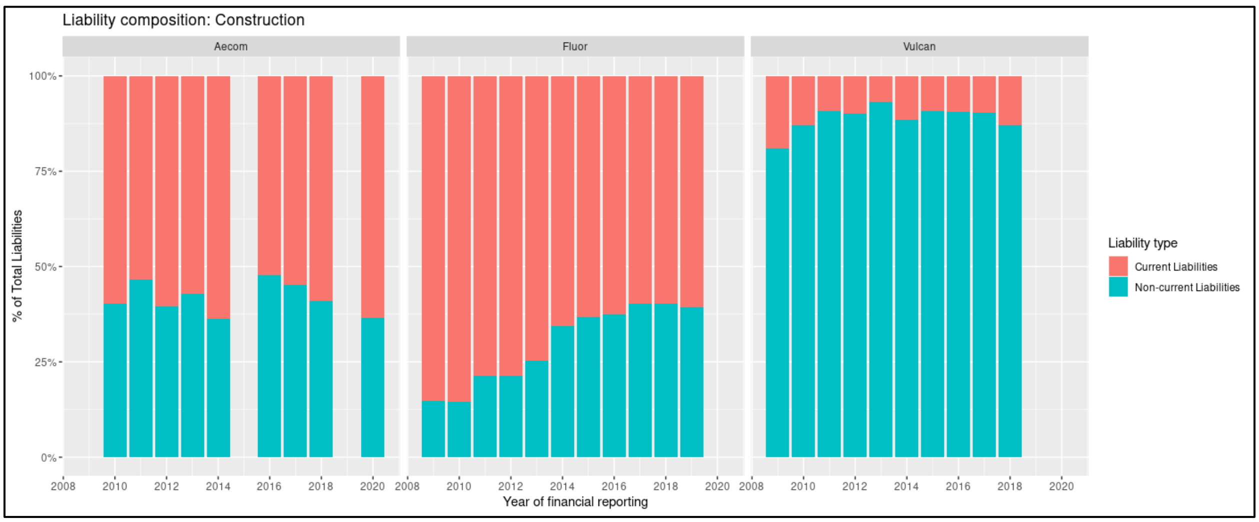
Automotive Liabilities



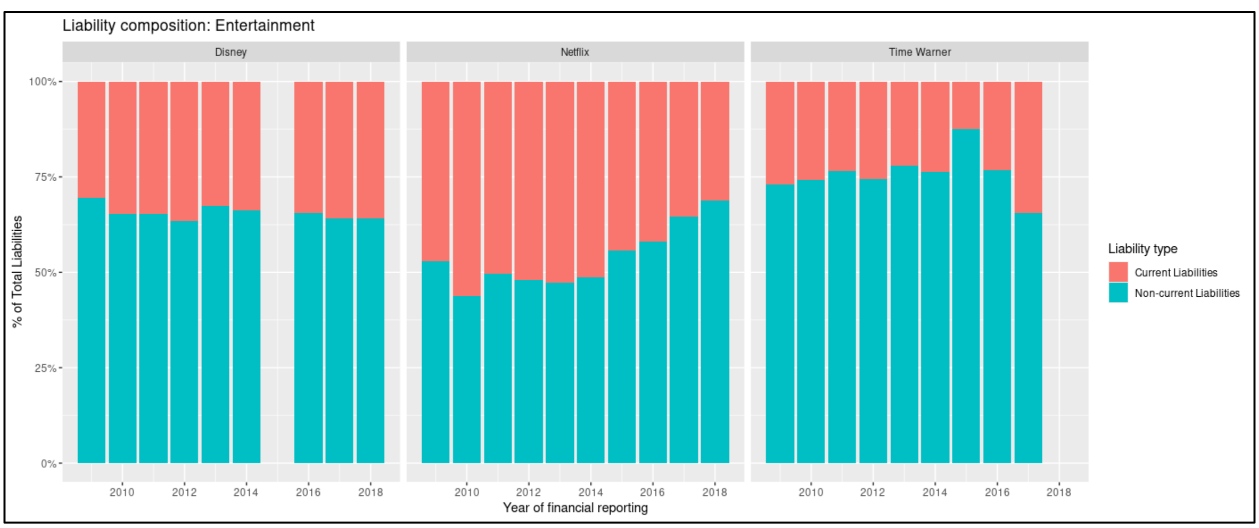
Biotech Liabilities



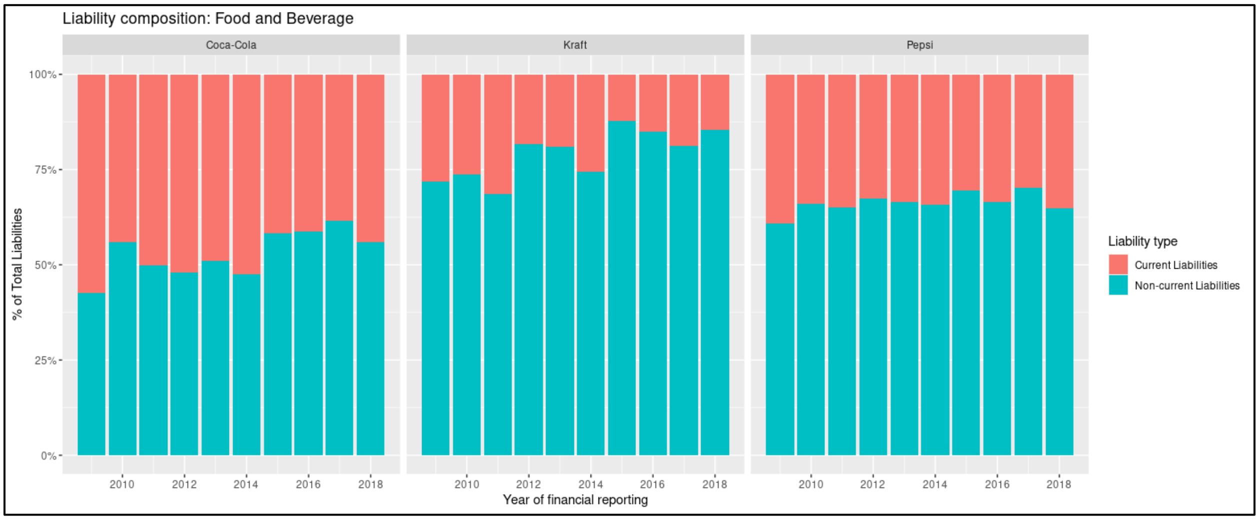
Construction Liabilities



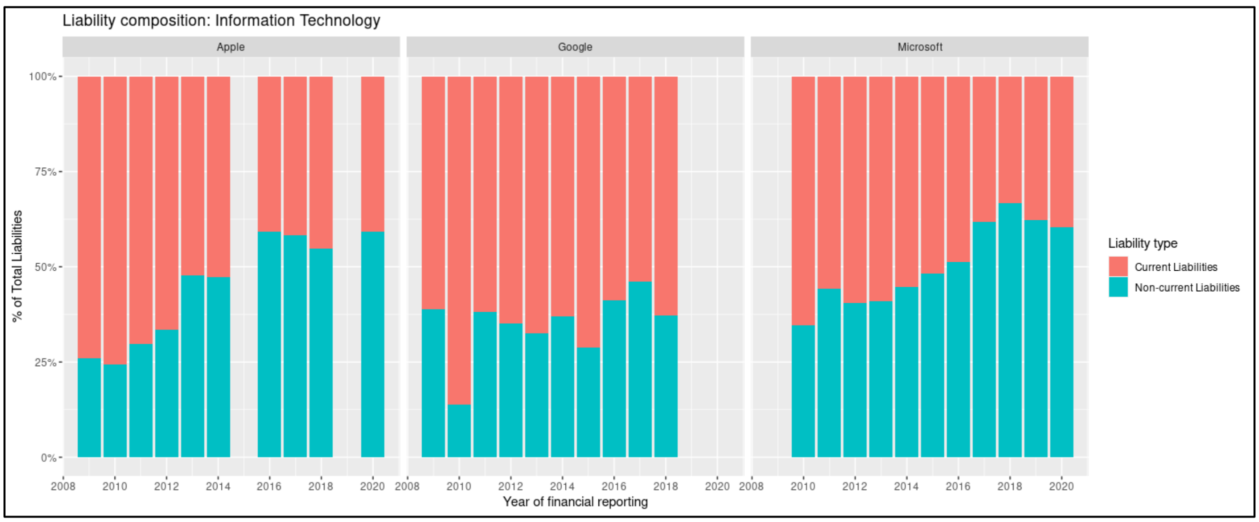
Entertainment Liabilities



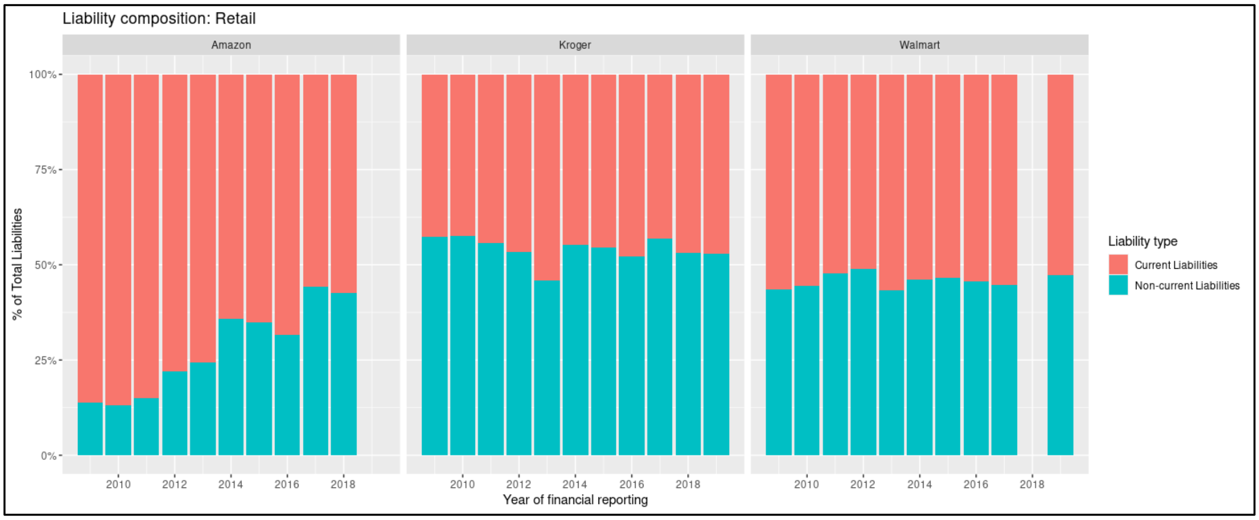
Food and Beverage Liabilities



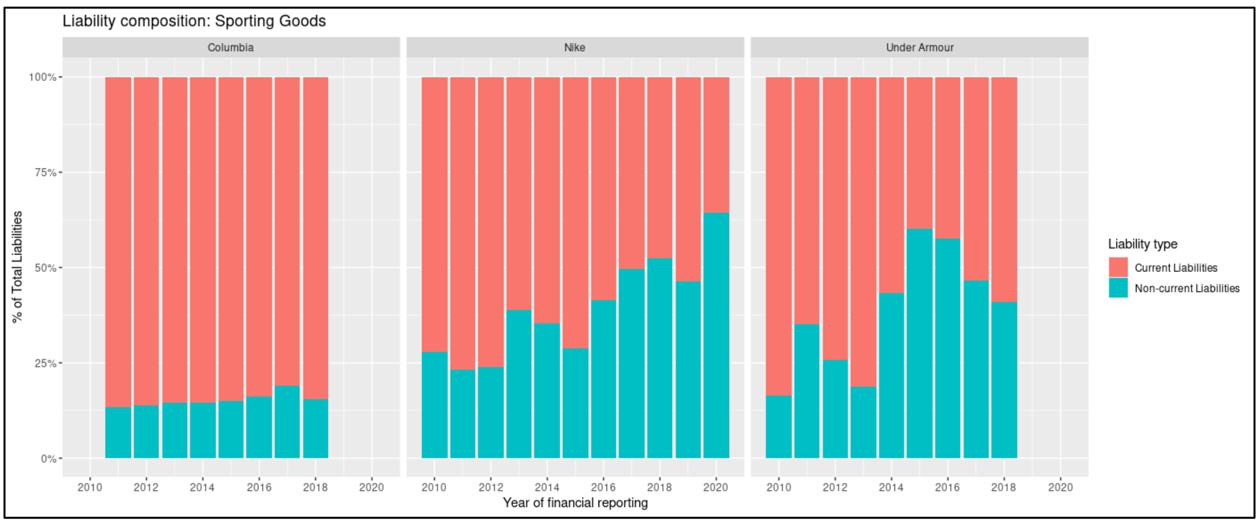
Information Technology Liabilities



Retail Liabilities

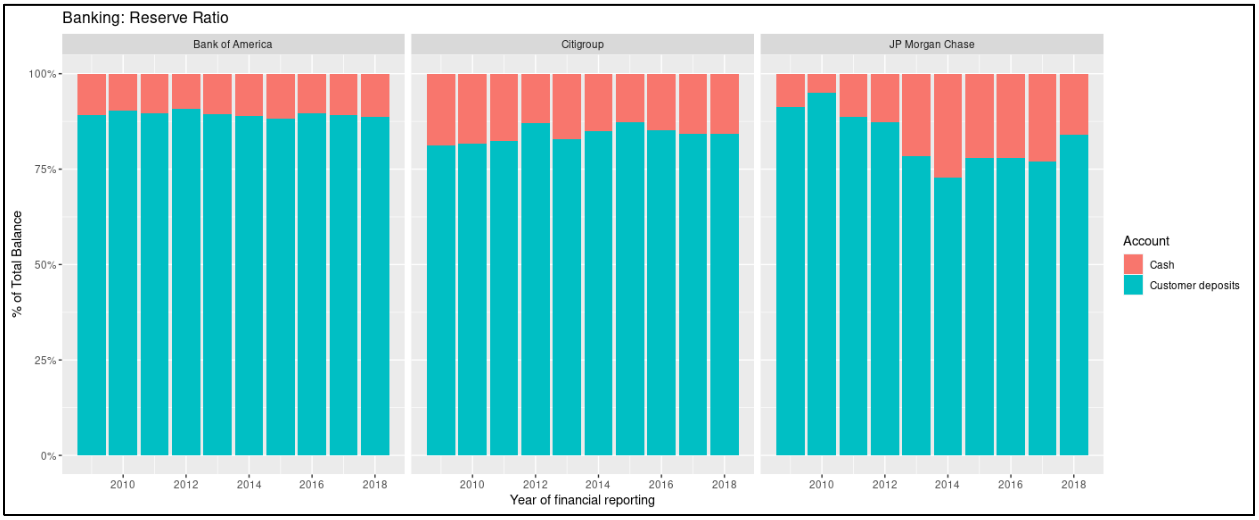


Sporting Goods Liabilities



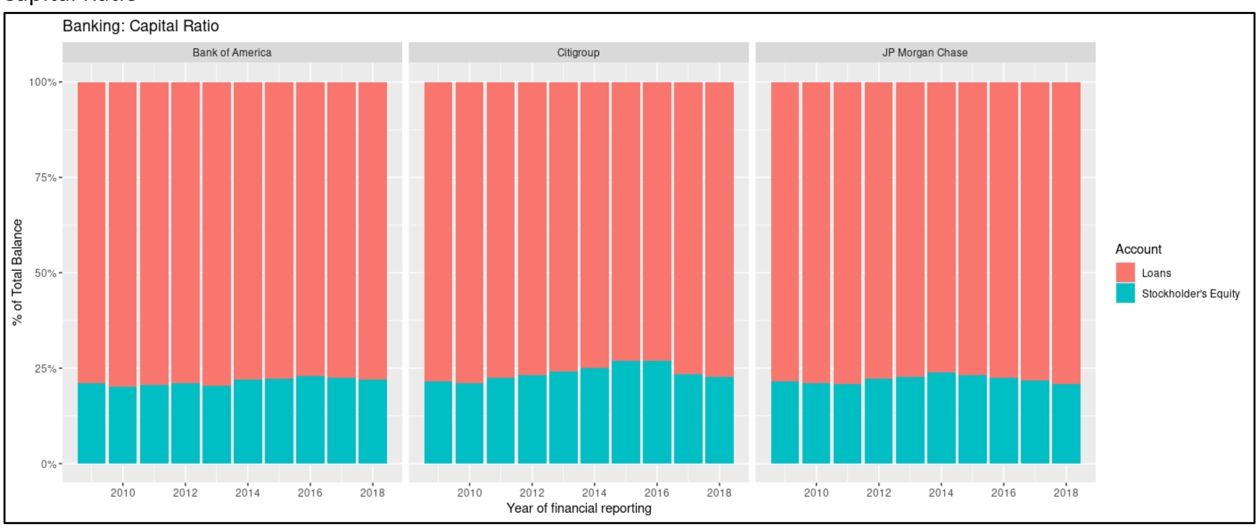
Bank Ratios <u>(for R code please see Appendix **E** at the end of the document)</u>

Reserve Ratio



A Reserve Ratio is a measure of how much cash a bank needs to hold onto instead of loaning out or investing and is a function of the amount of "Customer Deposits" a bank has. The commonly assumed requirement is 10% though almost no central bank and no major central bank imposes such a ratio requirement. As can be seen in the chart above, these three commercial banks do a fairly decent job of staying at or above 10%, with JP Morgan dipping below early on before having the highest ratio towards the end of the decade.

Capital Ratio

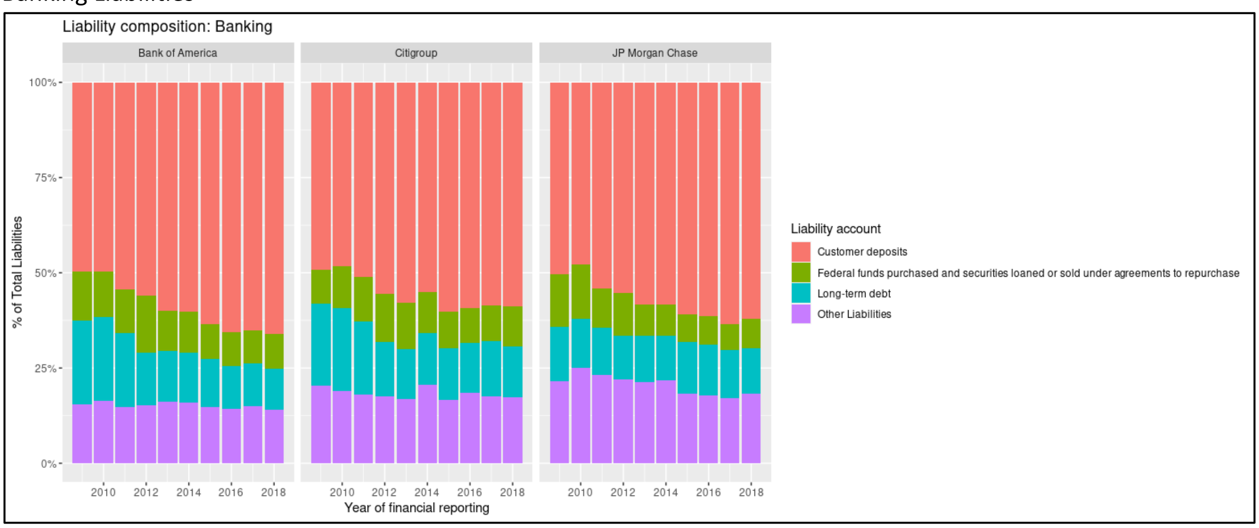


A Capital Ratio is essentially a bank's capital (or Equity) divided by the amount of Loans it has on the balance sheet. The idea here is that banks need to have a certain amount of capital compared to the amount of money they've loaned out to help offset the risk of those loans. Regulators typically track capital ratios to help assess the level of exposure a bank has to risk.

Banking Cash and cash equivalents

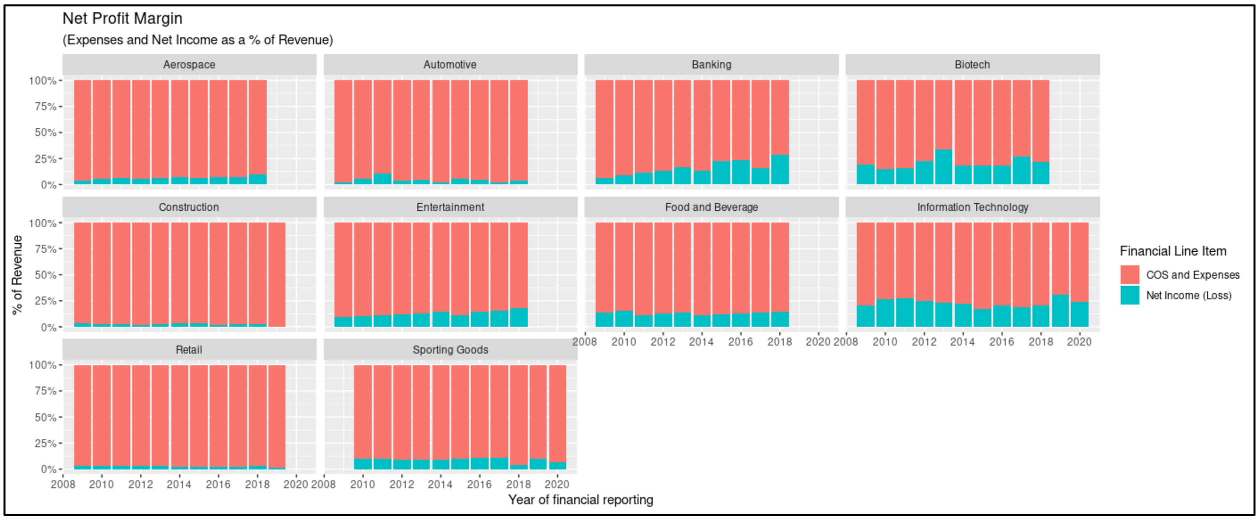
Banks interestingly have cash and cash equivalents divided into two main subgroups: "Cash and due from banks" and "Deposits with banks". The "Cash" from Cash and due from banks is literally the cash physically on hand at a bank's various locations and the "due from banks" is non interest earning cash owed from central banks. Deposits with banks on the other hand are the interest earning cash a commercial bank has deposited in central banks.

Banking Liabilities



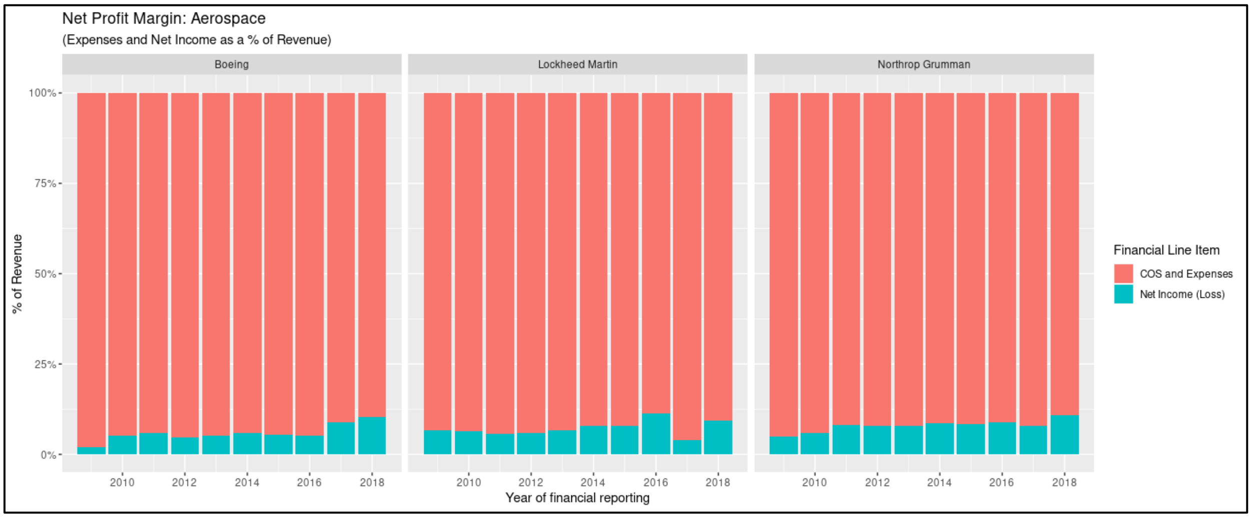
The above chart is a break down of the major liability accounts commercial banks have on their balance sheets. Really hits home to see that banks' number one liability is everyone else's money.

Net Profit Margin (for R code please see Appendix **E** at the end of the document)

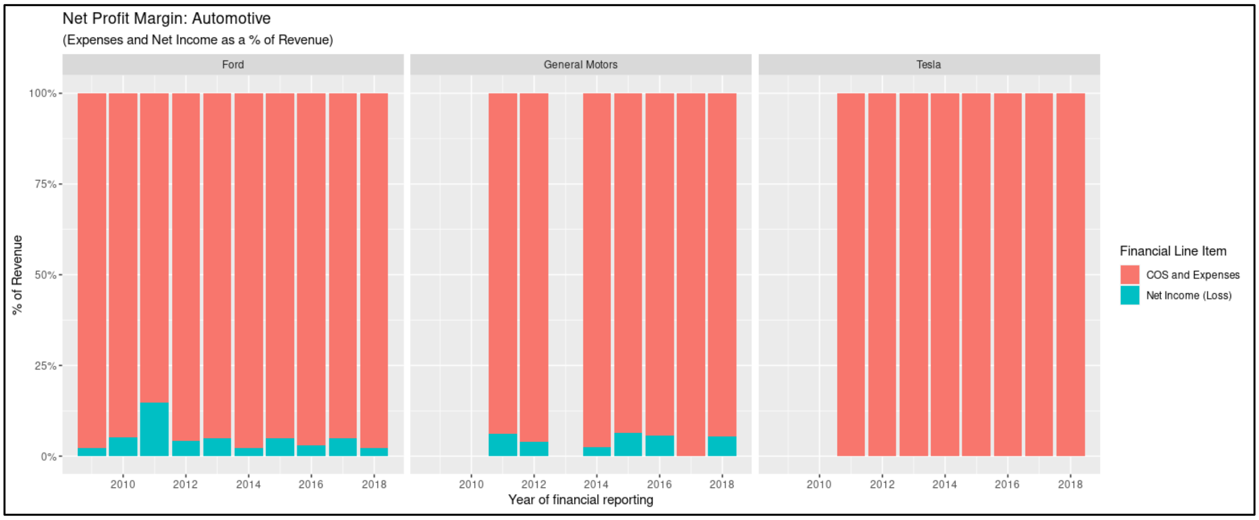


Similar to the A = L + E ratios, the chart above and the industry specific charts below are fairly straightforward but given how much of a staple Net Profit Margin is when it comes to financial ratios, generating stacked bar charts where Expenses and Net Income were displayed as percents of Revenue felt obligatory.

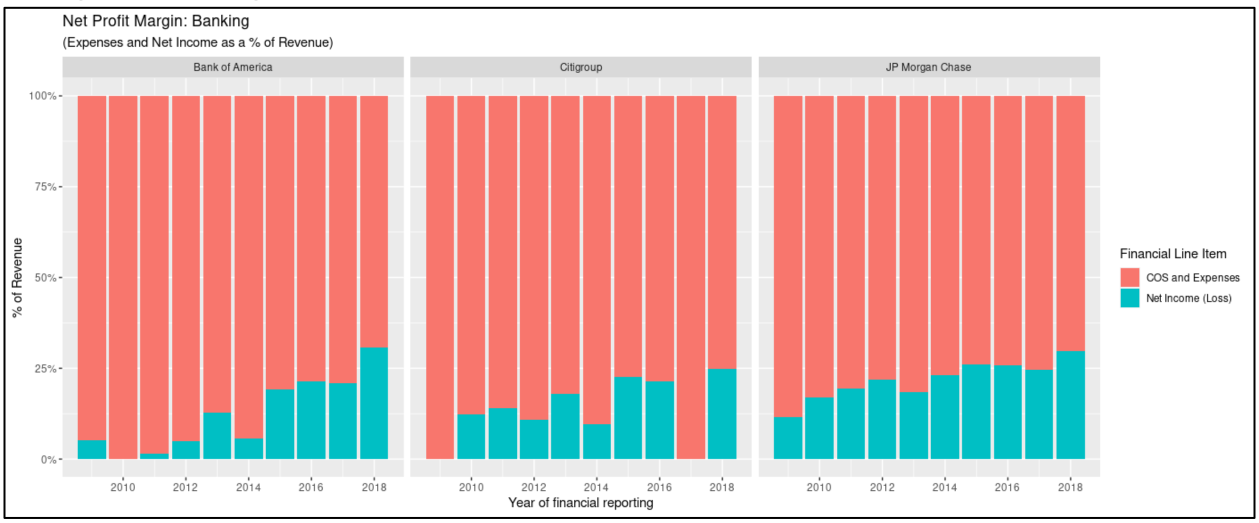
Aerospace Net Profit Margin



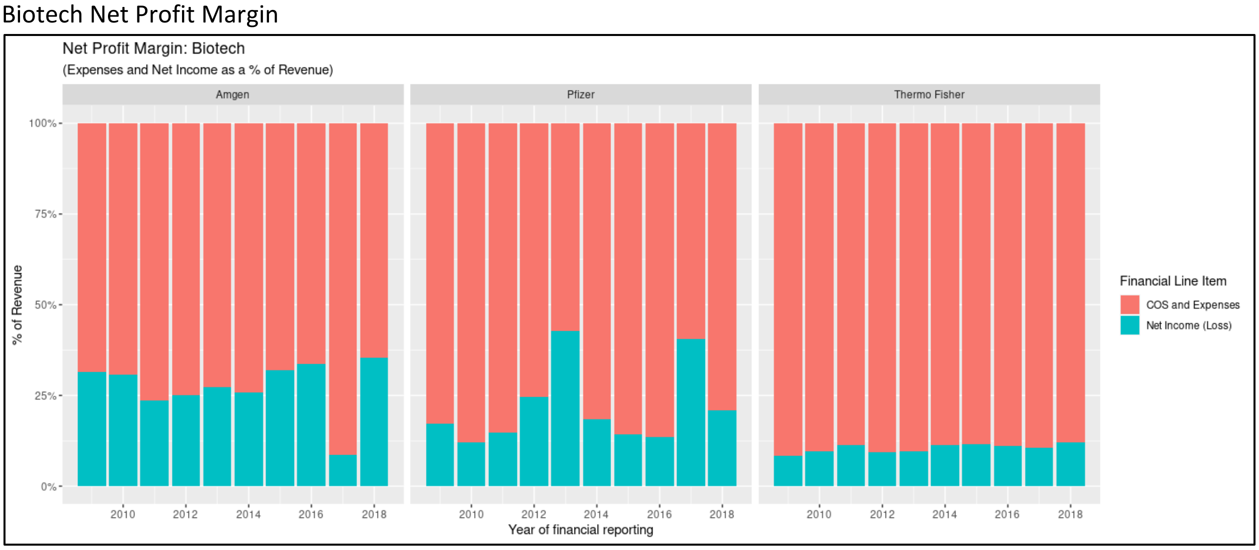
Automotive Net Profit Margin



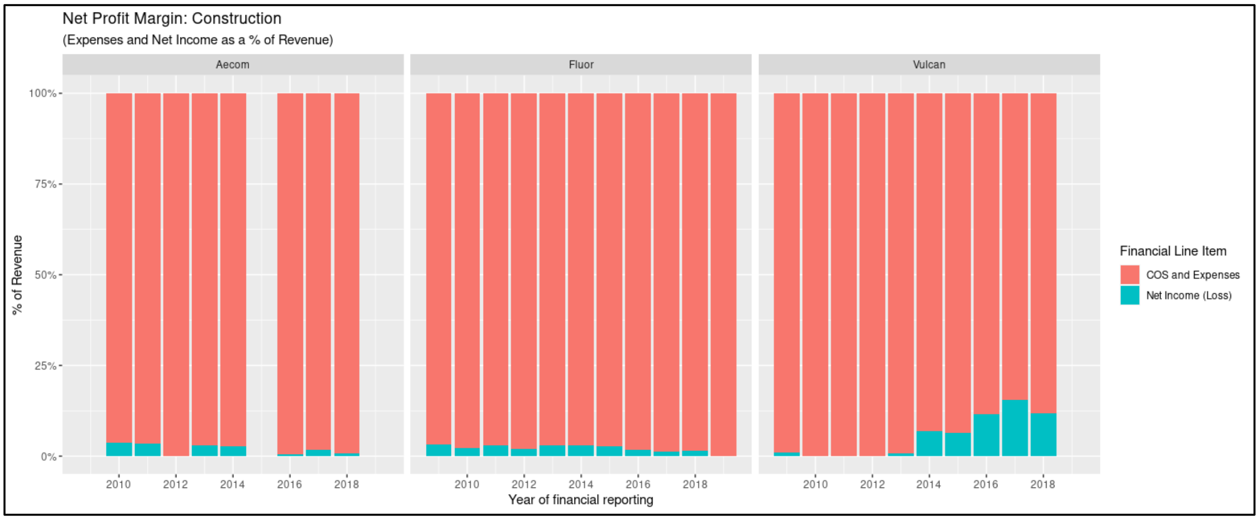
Banking Net Profit Margin



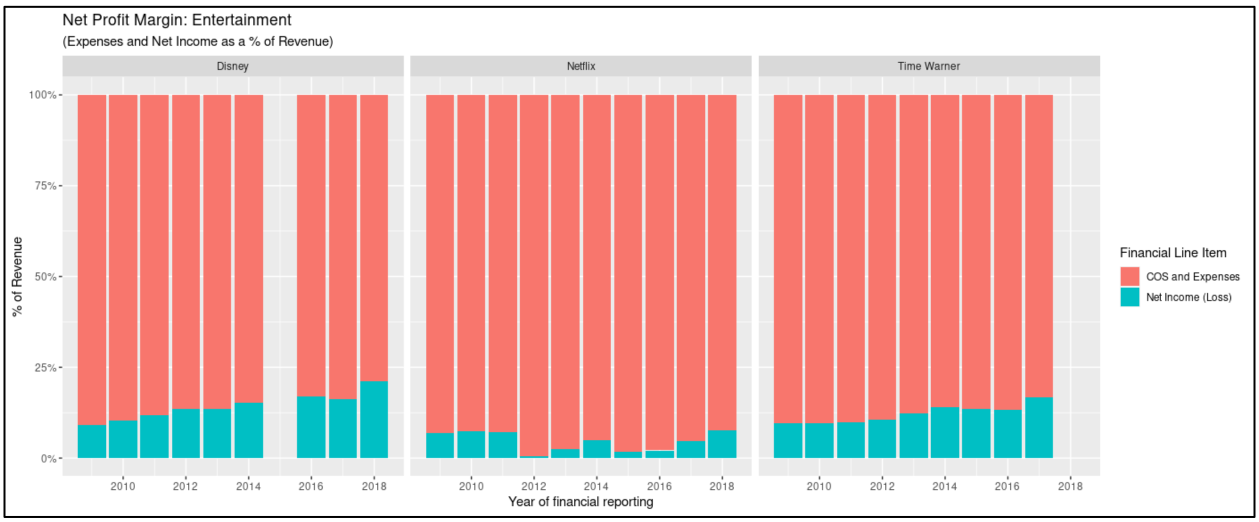
Biotech Net Profit Margin



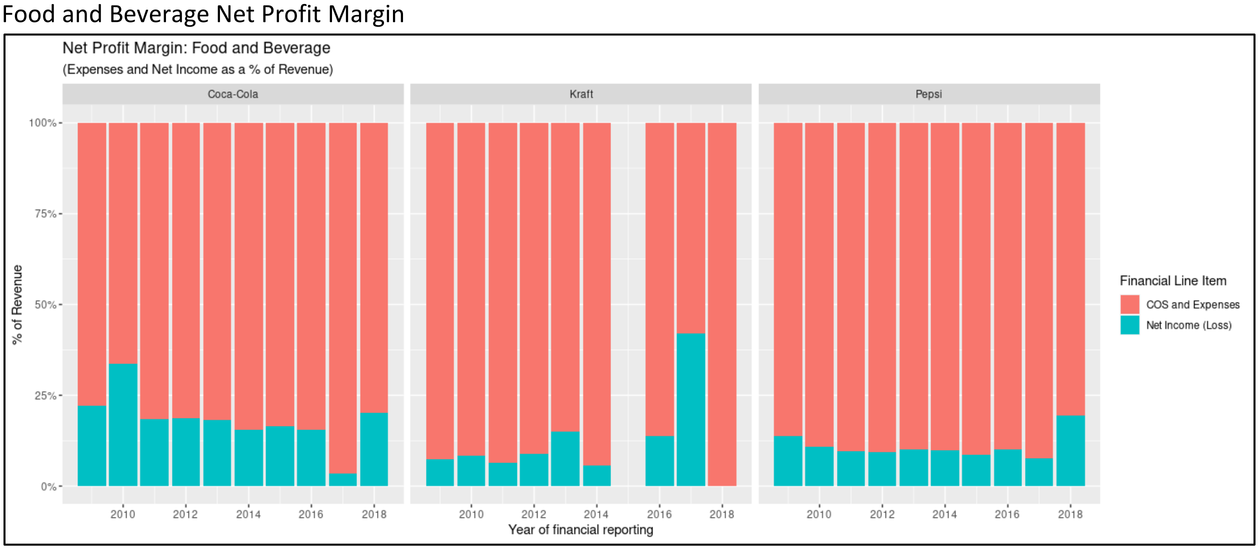
Construction Net Profit Margin



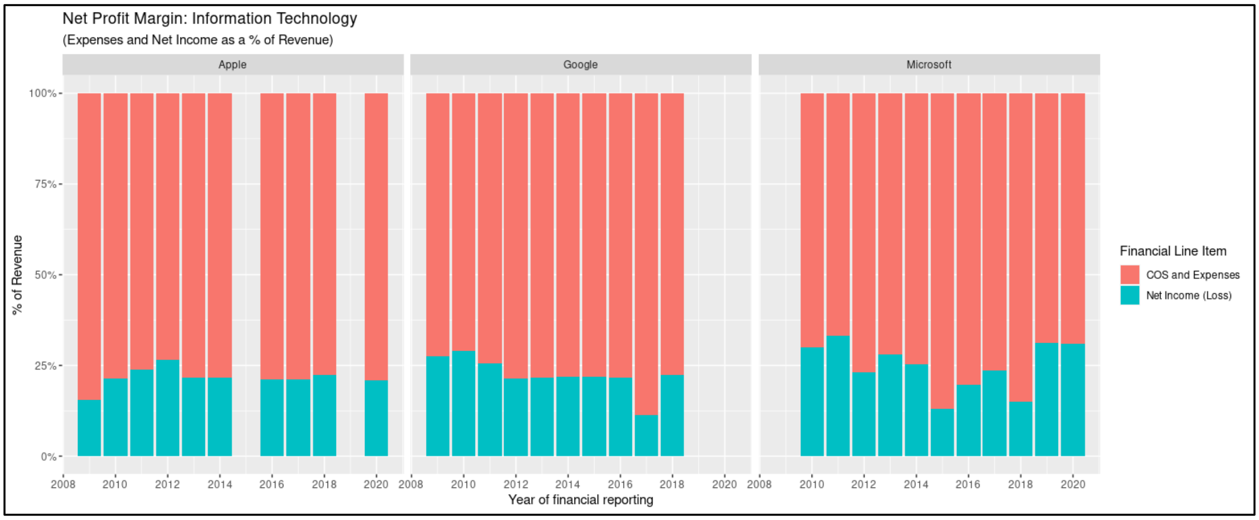
Entertainment Net Profit Margin



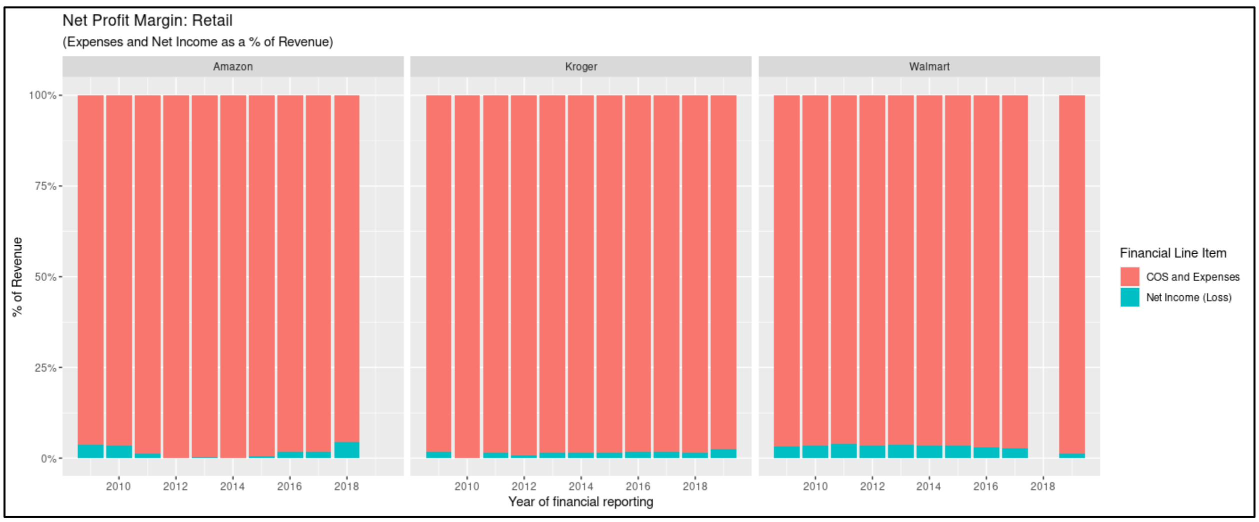
Food and Beverage Net Profit Margin



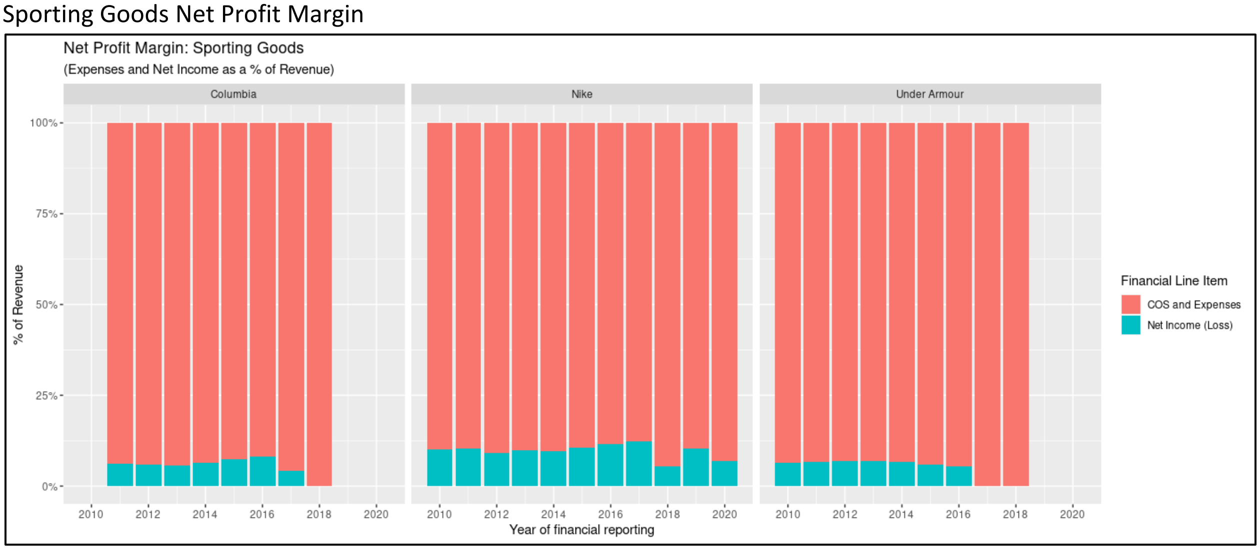
Information Technology Net Profit Margin



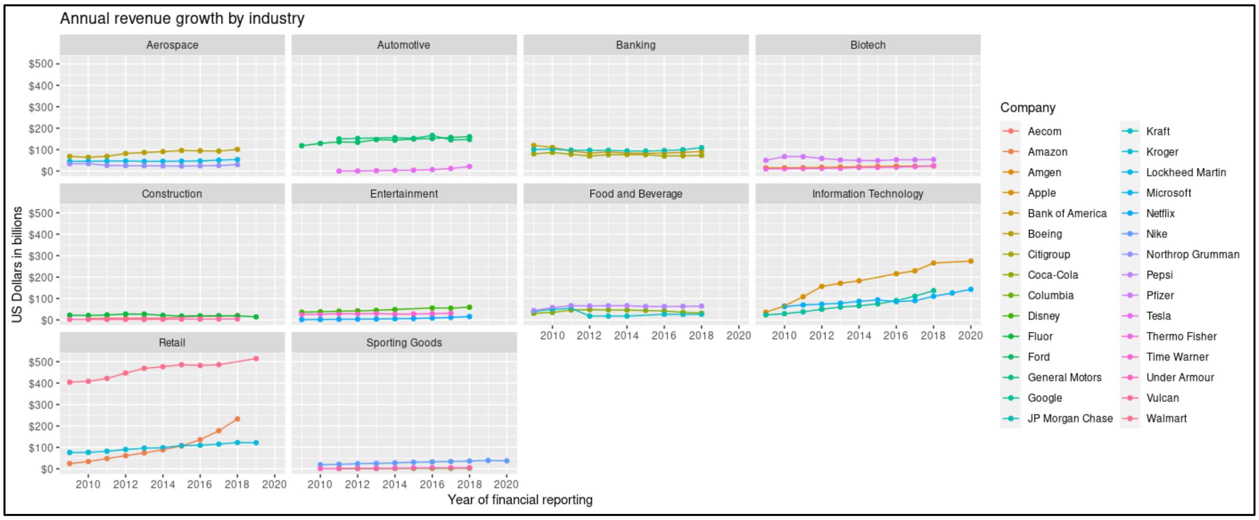
Retail Net Profit Margin



Sporting Goods Net Profit Margin



Annual Revenue Growth (for R code please see Appendix **B** at the end of the document)

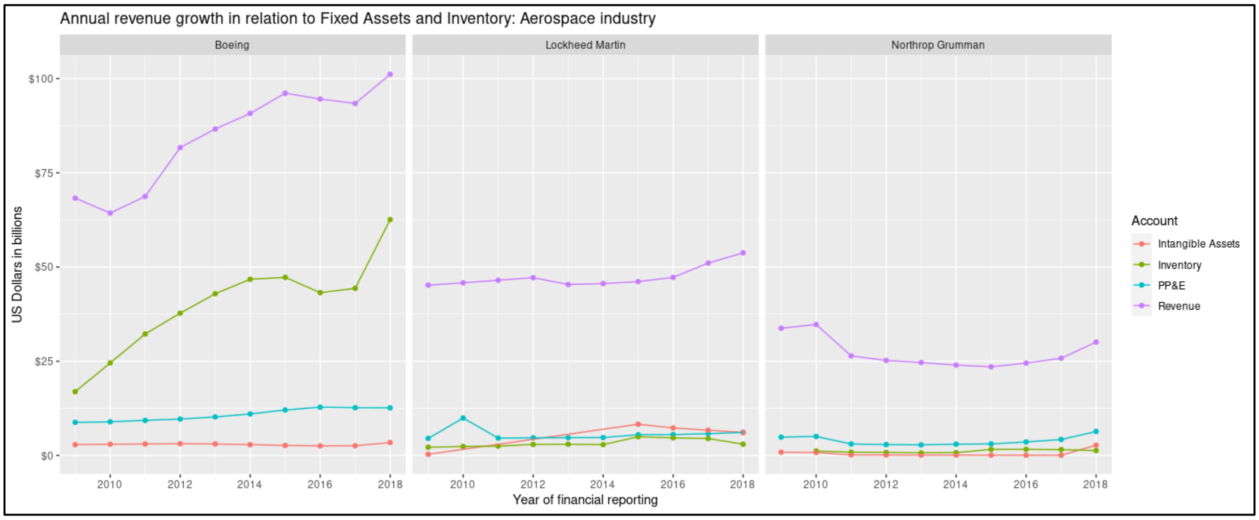


The plots above display Revenue growth over the last decade or so for each industry with the companies in each industry a different color. Walmart absolutely blows everyone out of the water, with only Apple and Amazon coming anywhere near half of Walmart's revenue. Industry wise, Retail, Information Technology, and Automotive tend to be the biggest performers, which, would not be unreasonable to expect.

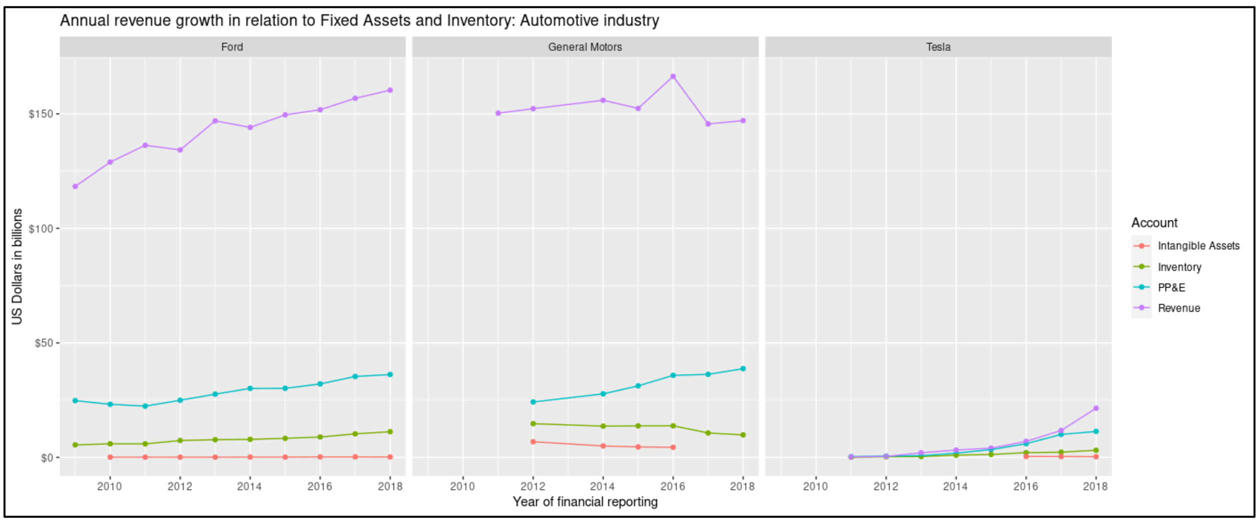
Annual Revenue Growth in relation to Fixed Assets and Inventory:

The final graphs in this project are plots of revenue for each company compared to the assets that actually produce revenue, such as inventory, PP&E, or intangible assets. As it might be expected, the asset classes that stood out in the Asset Composition section, for company or industry specific reasons, tend to be the assets that trend most with revenue. E.g., Boeing's inventory trends with its revenue, Disney's PP&E trends with its revenue, and intangible assets trend with the Biotech companies' revenue. Please find these plots on the three subsequent pages.

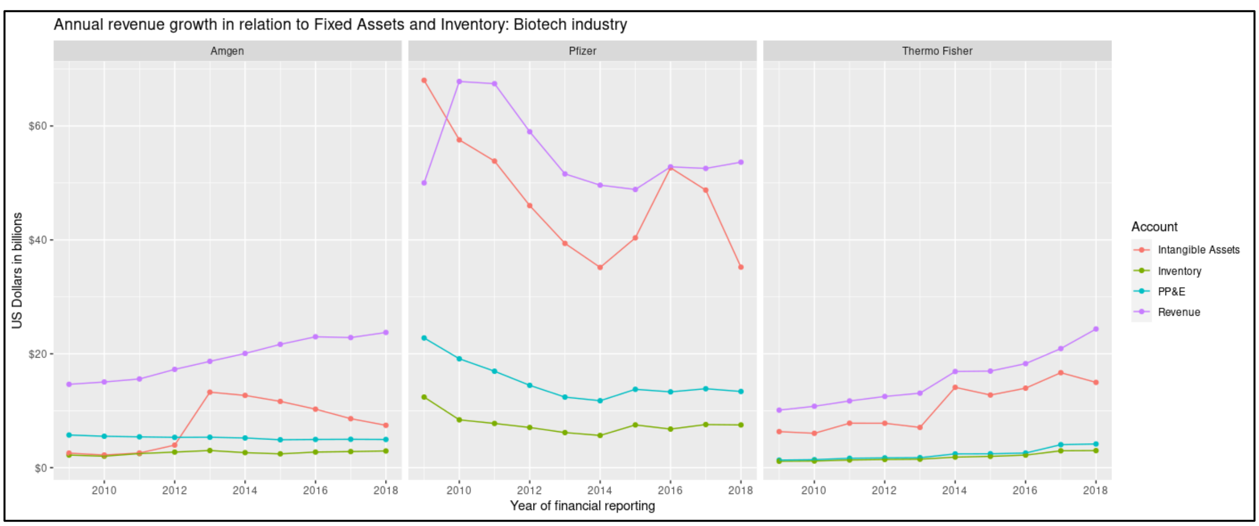
Aerospace Revenue



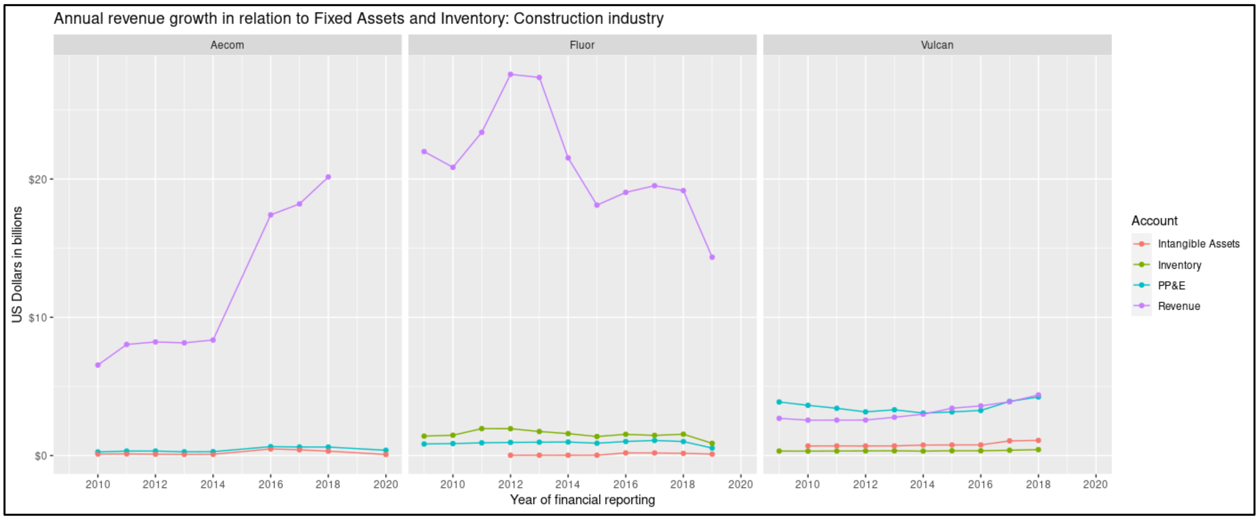
Automotive Revenue



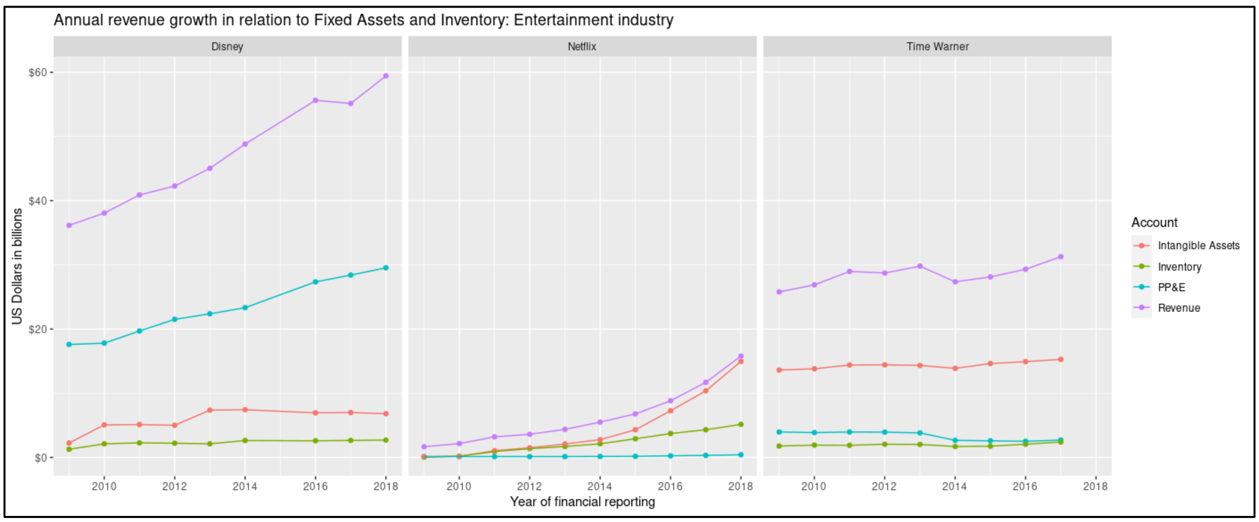
Biotech Revenue



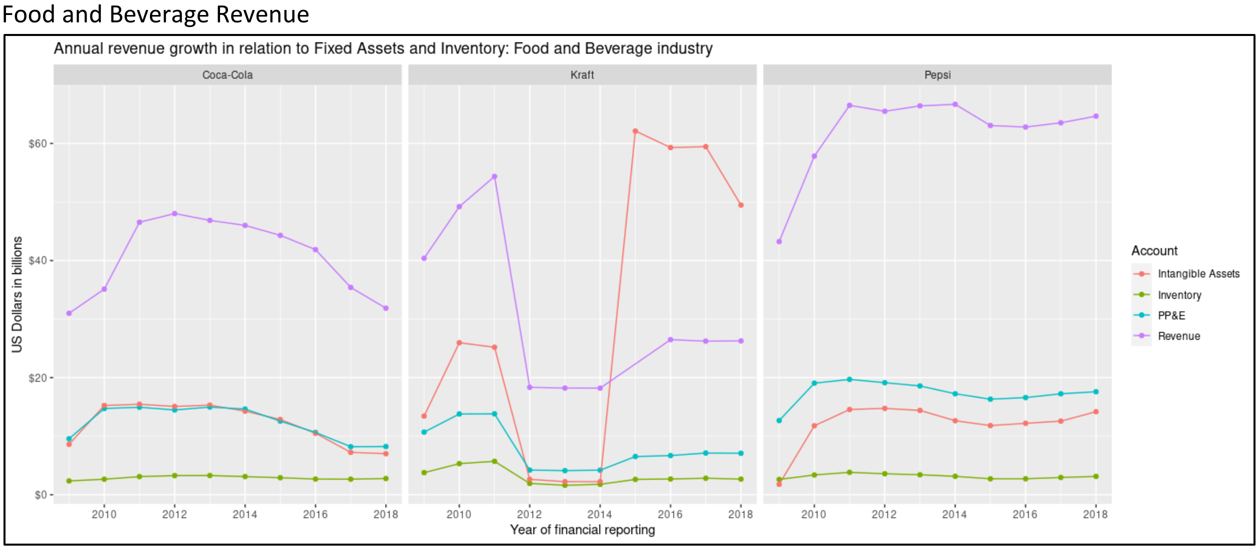
Construction Revenue



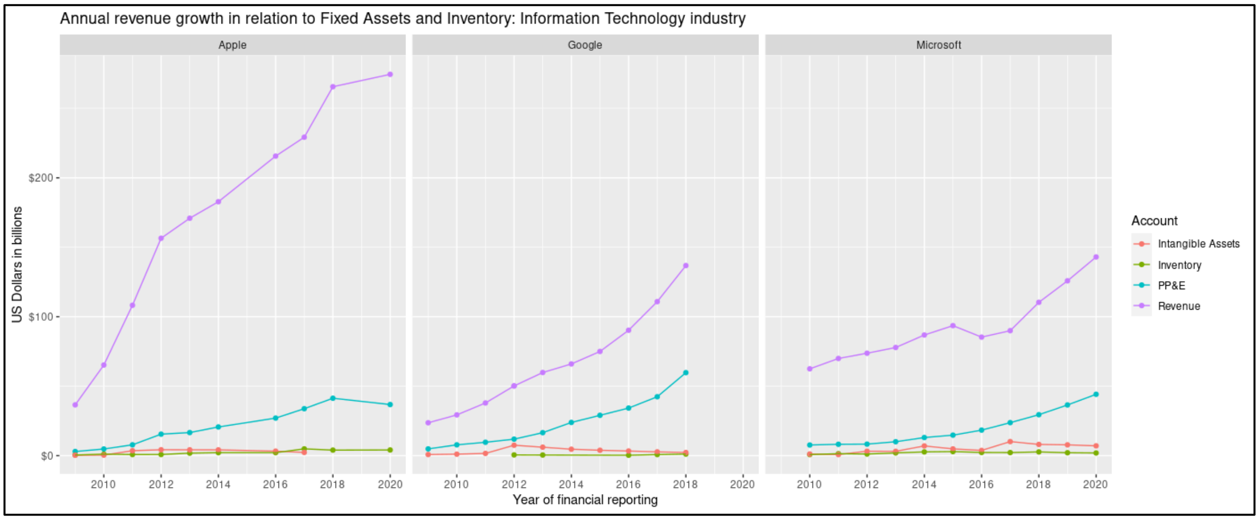
Entertainment Revenue



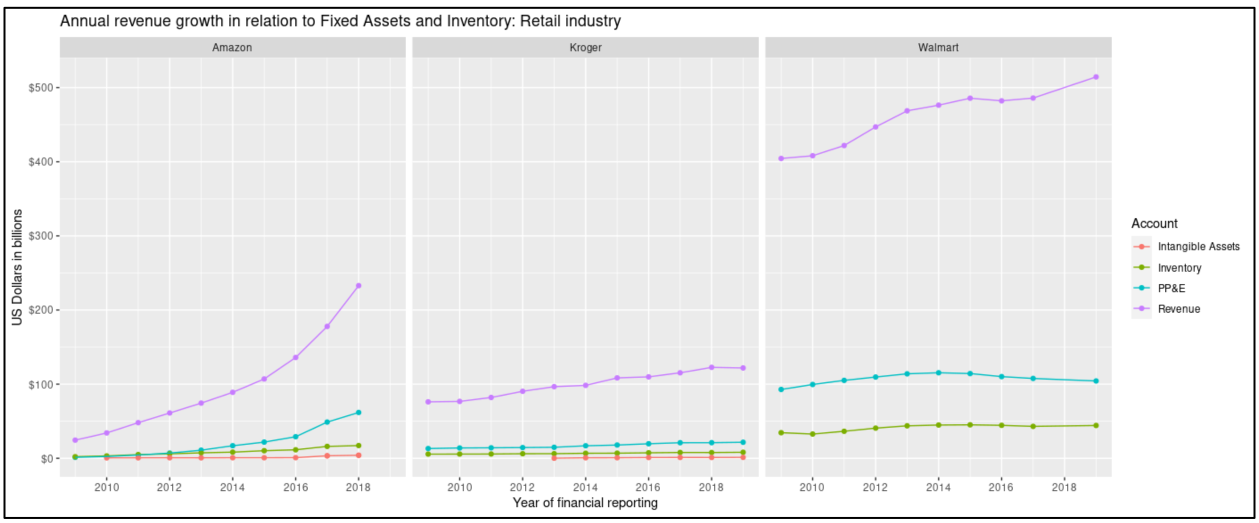
Food and Beverage Revenue



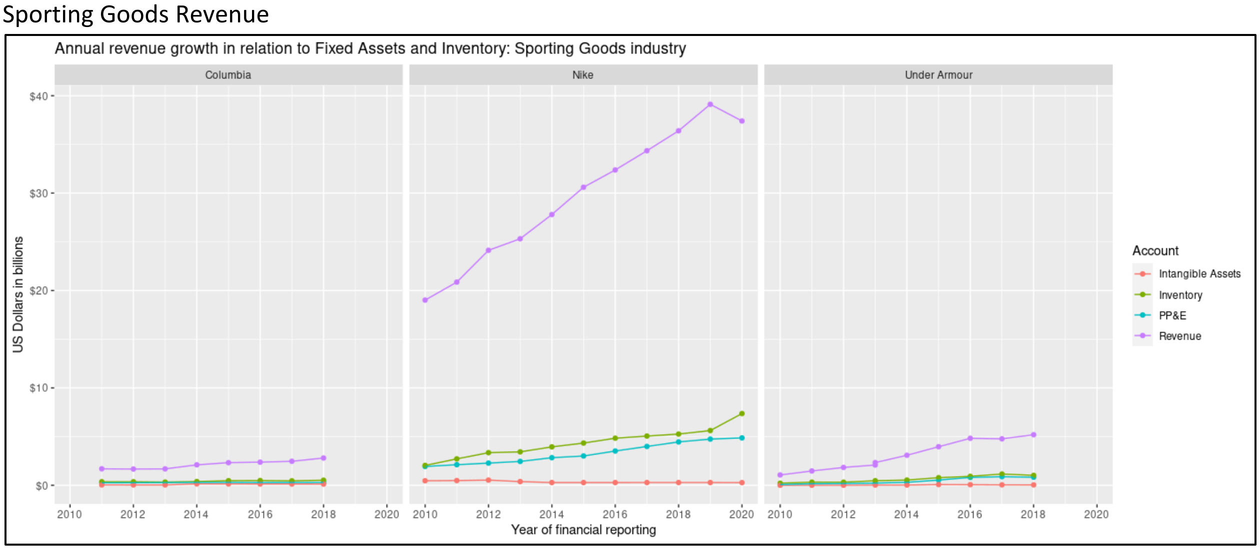
Information Technology Revenue



Retail Revenue



Sporting Goods Revenue



IV. Conclusion

I started this project over a year ago, the actual classes portion of the certificate program were fairly easy to complete, and I did so in an equally expedient manner. For the most part they were just review of what I had learned in school. This project I took my time with however, and when one has been working in public accounting like I have, spare time comes sparingly. So, I chipped away at what I could when I could. But now that it's done, I'm glad I had the time to dwell on and change things up as I am very pleased with the end result.

This project could have been a simple analysis of a simple dataset, but not for me. I wanted to tackle a dataset that hit home, and in a way that not only showcased my technical skills, but also my prowess when it comes to financial analysis and knowing my way around a balance sheet and income statement. And the result was actually a lot cooler than I thought it would be. The graphs and comparisons showed me things I didn't expect, affirmed somethings I did, and taught me a lot either way along the way.

Somethings I might want to explore in future analyses include but are not limited to - cash flow activity, how certain financial instruments behave in light of varying market conditions, or maybe something more macroscopic, like the credit cycle. I actually debated including items from the Statement of Cash Flows, but I figured I might dilute this project if I did, and I needed to get done with it eventually. I'll also need to learn a lot more if I want to tackle financial instruments or the credit cycle next time around.

In closing, it feels a bit like a small chapter of my life has closed, but in its stead a new library has been opened.

```
#Appendix A

library(tidyverse)
install.packages("tidyverse")

#all companies assets (excluding banking)
allcompaniesassetsnob=dplyr::filter(all_companies_v5, account_type == "Asset", industry !=
"Banking",
                                   account_name != "Current Assets" & account_name != "Non-current
Assets" &
                                   account_name !="Total Assets")
view(allcompaniesassetsnob)

ggplot(data=allcompaniesassetsnob, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition by industry",
    subtitle = "(Excluding Banking)", y = "% of Total Assets", x = "Year of financial reporting",
fill = "Asset account") + facet_wrap(~industry) +
  scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks = seq(2008,2020,by=2))

#banking assets
bankingassets=dplyr::filter(all_companies_v5, account_type == "Asset", industry == "Banking",
                            account_name != "Cash and due from banks" & account_name != "Deposits
with banks" & account_name !="Total Assets")
view(bankingassets)

ggplot(data=bankingassets, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition: Banking",
    subtitle = "(cash equivalents aggregated)", y = "% of Total Assets", x = "Year of financial
reporting", fill = "Asset account") + facet_wrap(~company_name) +
  scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks = seq(2008,2020,by=2))

bankingcash=dplyr::filter(all_companies_v5, account_type == "Asset", industry == "Banking",
                          account_name == "Cash and due from banks" | account_name == "Deposits
with banks")
view(bankingcash)

ggplot(data=bankingcash, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Banking:
Cash and cash equivalent composition",
   y = "% of Cash and cash equivalents", x = "Year of financial reporting", fill = "Cash account")
+ facet_wrap(~company_name) +
  scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks = seq(2008,2020,by=2))

#automotive assets
automotiveassets=dplyr::filter(all_companies_v5, account_type == "Asset", industry ==
"Automotive",
                               account_name != "Current Assets" & account_name != "Non-current
Assets" &
                                 account_name !="Total Assets")

ggplot(data=automotiveassets, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition: Automotive",
    y = "% of Total Assets", x = "Year of financial reporting", fill = "Asset account") +
facet_wrap(~company_name) +
  scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks = seq(2008,2020,by=2))

#aerospace assets
aerospaceassets=dplyr::filter(all_companies_v5, account_type == "Asset", industry == "Aerospace",
                               account_name != "Current Assets" & account_name != "Non-current
Assets" &
                                 account_name !="Total Assets")
```

```
view(aerospaceassets)

ggplot(data=aerospaceassets, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition: Aerospace",
   y = "% of Total Assets", x = "Year of financial reporting", fill = "Asset account") +
facet_wrap(~company_name) +
  scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks = seq(2008,2020,by=2))

#information technology assets
ITassets=dplyr::filter(all_companies_v5, account_type == "Asset", industry == "Information
Technology",
                               account_name != "Current Assets" & account_name != "Non-current
Assets" &
                                 account_name !="Total Assets")

ggplot(data=ITassets, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition: Information Technology",
    y = "% of Total Assets", x = "Year of financial reporting", fill = "Asset account") +
facet_wrap(~company_name) +
  scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks = seq(2008,2020,by=2))

#construction assets
constassets=dplyr::filter(all_companies_v5, account_type == "Asset", industry == "Construction",
                       account_name != "Current Assets" & account_name != "Non-current Assets" &
                         account_name !="Total Assets")

ggplot(data=constassets, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition: Construction",
   y = "% of Total Assets", x = "Year of financial reporting", fill = "Asset account") +
facet_wrap(~company_name) +
  scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks = seq(2008,2020,by=2))

#retail assets
retailassets=dplyr::filter(all_companies_v5, account_type == "Asset", industry == "Retail",
                          account_name != "Current Assets" & account_name != "Non-current Assets"
&
                            account_name !="Total Assets")

ggplot(data=retailassets, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition: Retail",
   y = "% of Total Assets", x = "Year of financial reporting", fill = "Asset account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#biotech assets
biotechassets=dplyr::filter(all_companies_v5, account_type == "Asset", industry == "Biotech",
                           account_name != "Current Assets" & account_name != "Non-current Assets"
&
                             account_name !="Total Assets")

ggplot(data=biotechassets, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition: Biotech",
   y = "% of Total Assets", x = "Year of financial reporting", fill = "Asset account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#entertainment assets
enterassets=dplyr::filter(all_companies_v5, account_type == "Asset", industry == "Entertainment",
```

```
                            account_name != "Current Assets" & account_name != "Non-current
Assets" &
                              account_name !="Total Assets")

ggplot(data=enterassets, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition: Entertainment",
    y = "% of Total Assets", x = "Year of financial reporting", fill = "Asset account") +
facet_wrap(~company_name) +
      scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#sport assets
sportassets=dplyr::filter(all_companies_v5, account_type == "Asset", industry == "Sporting Goods",
                          account_name != "Current Assets" & account_name != "Non-current Assets"
&
                            account_name !="Total Assets")

ggplot(data=sportassets, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition: Sporting Goods",
   y = "% of Total Assets", x = "Year of financial reporting", fill = "Asset account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#food and beverage assets
foodassets=dplyr::filter(all_companies_v5, account_type == "Asset", industry == "Food and
Beverage",
                          account_name != "Current Assets" & account_name != "Non-current Assets"
&
                            account_name !="Total Assets")

ggplot(data=foodassets, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Asset
composition: Food and Beverage",
   y = "% of Total Assets", x = "Year of financial reporting", fill = "Asset account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))
```

```
#Appendix B

#all companies
company_revenue <- dplyr::filter(all_companies_v5, account_name == "Revenue")

ggplot(data=company_revenue, aes(x = fiscal_year, y = usd_in_000s, color = company_name)) +
  geom_point() + geom_line() + facet_wrap(~industry) + labs(title = "Annual revenue growth by
industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Company") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

view(company_revenue)

#aerospace
aerospace_revenue <- dplyr::filter(company_revenue, industry == "Aerospace")
aerospace_revenuefainv <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name
== "PP&E" |
                                       account_name == "Intangible Assets" | account_name ==
"Inventory", industry == "Aerospace")
aerospace_revenuecashar <- dplyr::filter(all_companies_v5, account_name == "Revenue" |
account_name == "Cash" |
                                           account_name == "Accounts Receivable", industry ==
"Aerospace")

view(aerospace_revenue)

ggplot(data=aerospace_revenuefainv, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Fixed Assets and Inventory: Aerospace industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

ggplot(data=aerospace_revenuecashar, aes(x = fiscal_year, y = usd_in_000s, color = account_name))
+
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Cash and Accounts Receivable: Aerospace industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#automotive
automotive_revenue <- dplyr::filter(all_companies_v5, account_name == "Revenue", industry ==
"Automotive")
auto_revenuefainv <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"PP&E" |
                                       account_name == "Intangible Assets" | account_name ==
"Inventory", industry == "Automotive")
auto_revenuecashar <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"Cash" |
                                           account_name == "Accounts Receivable", industry ==
"Automotive")

ggplot(data=auto_revenuefainv, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Fixed Assets and Inventory: Automotive industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

ggplot(data=auto_revenuecashar, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Cash and Accounts Receivable: Automotive industry",
```

```
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#banking
banking_revenue <- dplyr::filter(all_companies_v5, account_name == "Revenue", industry ==
"Banking")
banking_revenueloan <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"Loans", industry == "Banking")
banking_revenueall <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"Loans" | account_name == "Investment Securities" |
                                      account_name == "Trading account assets" | account_name ==
"Federal funds sold and securities borrowed and purchased under agreements to resell", industry ==
"Banking")

ggplot(data=banking_revenueall, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual Banking revenue
growth in relation to interest earning assets and assets held for sell",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#biotech
biotech_revenue <- dplyr::filter(all_companies_v5, account_name == "Revenue", industry ==
"Biotech")
bio_revenuefainv <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"PP&E" |
                                     account_name == "Intangible Assets" | account_name ==
"Inventory", industry == "Biotech")
bio_revenuecashar <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"Cash" |
                                      account_name == "Accounts Receivable", industry ==
"Biotech")

ggplot(data=bio_revenuefainv, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Fixed Assets and Inventory: Biotech industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

ggplot(data=bio_revenuecashar, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Cash and Accounts Receivable: Biotech industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#construction
construction_revenue <- dplyr::filter(all_companies_v5, account_name == "Revenue", industry ==
"Construction")
const_revenuefainv <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"PP&E" |
                                    account_name == "Intangible Assets" | account_name ==
"Inventory", industry == "Construction")
const_revenuecashar <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"Cash" |
                                     account_name == "Accounts Receivable", industry ==
"Construction")

ggplot(data=const_revenuefainv, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Fixed Assets and Inventory: Construction industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
```

```
seq(2008,2020,by=2))

ggplot(data=const_revenuecashar, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Cash and Accounts Receivable: Construction industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#entertainment
entertainment_revenue <- dplyr::filter(all_companies_v5, account_name == "Revenue", industry ==
"Entertainment")
entertainment_revenuefainv <- dplyr::filter(all_companies_v5, account_name == "Revenue" |
account_name == "PP&E" |
                                      account_name == "Intangible Assets" | account_name ==
"Inventory", industry == "Entertainment")
entertainment_revenuecashar <- dplyr::filter(all_companies_v5, account_name == "Revenue" |
account_name == "Cash" |
                                       account_name == "Accounts Receivable", industry ==
"Entertainment")

ggplot(data=entertainment_revenuefainv, aes(x = fiscal_year, y = usd_in_000s, color =
account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Fixed Assets and Inventory: Entertainment industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

ggplot(data=entertainment_revenuecashar, aes(x = fiscal_year, y = usd_in_000s, color =
account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Cash and Accounts Receivable: Entertainment industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#food and beverage
food_revenue <- dplyr::filter(all_companies_v5, account_name == "Revenue", industry == "Food and
Beverage")
food_revenuefainv <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"PP&E" |
                                       account_name == "Intangible Assets" | account_name ==
"Inventory", industry == "Food and Beverage")
food_revenuecashar <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"Cash" |
                                        account_name == "Accounts Receivable", industry == "Food
and Beverage")

ggplot(data=food_revenuefainv, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Fixed Assets and Inventory: Food and Beverage industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

ggplot(data=food_revenuecashar, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Cash and Accounts Receivable: Food and Beverage industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#information technology
info_revenue <- dplyr::filter(all_companies_v5, account_name == "Revenue", industry ==
```

```
"Information Technology")
view(info_revenue)

info_revenuefainv <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"PP&E" |
                                    account_name == "Intangible Assets" | account_name ==
"Inventory", industry == "Information Technology")
info_revenuecashar <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"Cash" |
                                     account_name == "Accounts Receivable", industry ==
"Information Technology")

ggplot(data=info_revenuefainv, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Fixed Assets and Inventory: Information Technology industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

ggplot(data=info_revenuecashar, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Cash and Accounts Receivable: Information Technology industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#retail
retail_revenue <- dplyr::filter(all_companies_v5, account_name == "Revenue", industry == "Retail")
retail_revenuefainv <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"PP&E" |
                                  account_name == "Intangible Assets" | account_name ==
"Inventory", industry == "Retail")
retail_revenuecashar <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name
== "Cash" |
                                  account_name == "Accounts Receivable", industry == "Retail")

view(retail_revenue)

ggplot(data=retail_revenuefainv, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Fixed Assets and Inventory: Retail industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

ggplot(data=retail_revenuecashar, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Cash and Accounts Receivable: Retail industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))



#sporting goods
sportinggoods_revenue <- dplyr::filter(all_companies_v5, account_name == "Revenue", industry ==
"Sporting Goods")
sport_revenuefainv <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"PP&E" |
                                      account_name == "Intangible Assets" | account_name ==
"Inventory", industry == "Sporting Goods")
sport_revenuecashar <- dplyr::filter(all_companies_v5, account_name == "Revenue" | account_name ==
"Cash" |
                                       account_name == "Accounts Receivable", industry ==
"Sporting Goods")
```

```
ggplot(data=sport_revenuefainv, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Fixed Assets and Inventory: Sporting Goods industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

ggplot(data=sport_revenuecashar, aes(x = fiscal_year, y = usd_in_000s, color = account_name)) +
  geom_point() + geom_line() + facet_wrap(~company_name) + labs(title = "Annual revenue growth in
relation to Cash and Accounts Receivable: Sporting Goods industry",
   y = "US Dollars in billions", x = "Year of financial reporting", color = "Account") +
    scale_y_continuous(labels = scales::dollar_format(scale = 1e-6)) + scale_x_continuous(breaks =
seq(2008,2020,by=2))
```

```
#Appendix C

#all companies
allcompaniesale=dplyr::filter(all_companies_v5, usd_in_000s > 0, account_name == "Total Assets" |
                                account_name == "Total Liabilities" | account_name
=="Stockholder's Equity")
view(allcompaniesale)

ggplot(data=allcompaniesale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
   subtitle = "All Industries", y = "% of Total", x = "Year of financial reporting", fill =
"Account") +
    facet_wrap(~industry) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#aerospace
aeroale=dplyr::filter(allcompaniesale, industry == "Aerospace")

ggplot(data=aeroale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
    subtitle = "Industry: Aerospace",  y = "% of Total", x = "Year of financial reporting", fill =
"Account") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#automotive
autoale=dplyr::filter(allcompaniesale, industry == "Automotive")

ggplot(data=autoale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
    subtitle = "Industry: Automotive",  y = "% of Total", x = "Year of financial reporting", fill
= "Account") +
      facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#banking
bankale=dplyr::filter(allcompaniesale, industry == "Banking")

ggplot(data=bankale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
   subtitle = "Industry: Banking",  y = "% of Total", x = "Year of financial reporting", fill =
"Account") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#biotech
bioale=dplyr::filter(allcompaniesale, industry == "Biotech")

ggplot(data=bioale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
    subtitle = "Industry: Biotech",  y = "% of Total", x = "Year of financial reporting", fill =
"Account") +
      facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#construction
constale=dplyr::filter(allcompaniesale, industry == "Construction")

ggplot(data=constale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
```

```
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
    subtitle = "Industry: Construction",  y = "% of Total", x = "Year of financial reporting",
fill = "Account") +
      facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#entertainment
enterale=dplyr::filter(allcompaniesale, industry == "Entertainment")

ggplot(data=enterale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
   subtitle = "Industry: Entertainment",  y = "% of Total", x = "Year of financial reporting",
fill = "Account") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#foodandbeverage
foodale=dplyr::filter(allcompaniesale, industry == "Food and Beverage")

ggplot(data=foodale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
    subtitle = "Industry: Food and Beverage",  y = "% of Total", x = "Year of financial
reporting", fill = "Account") +
      facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#informationtechnology
infoale=dplyr::filter(allcompaniesale, industry == "Information Technology")

ggplot(data=infoale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
   subtitle = "Industry: Information Technology",  y = "% of Total", x = "Year of financial
reporting", fill = "Account") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#retail
retailale=dplyr::filter(allcompaniesale, industry == "Retail")

ggplot(data=retailale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
   subtitle = "Industry: Retail",  y = "% of Total", x = "Year of financial reporting", fill =
"Account") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#sport
sportale=dplyr::filter(allcompaniesale, industry == "Sporting Goods")

ggplot(data=sportale, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Total
Assets, Total Liabilities, and Stockholder's Equity",
   subtitle = "Industry: Sporting Goods",  y = "% of Total", x = "Year of financial reporting",
fill = "Account") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))
```

```
#Appendix D

#current ratios
#all companies
allcompaniescurrent <- dplyr::filter(all_companies_v5, industry != "Banking",
                                     account_name == "Current Assets" |
                                       account_name == "Current Liabilities")
view(allcompaniescurrent)

ggplot(data=allcompaniescurrent, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Current
Ratio by industry",
    subtitle = "(Excluding Banking)", y = "% of Total Balance", x = "Year of financial reporting",
fill = "Asset account") +
      facet_wrap(~industry) + scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))

#aerospace
aerocurrent <- dplyr::filter(allcompaniescurrent, industry == "Aerospace")

ggplot(data=aerocurrent, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Current
Ratio: Aerospace",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#automotive
autocurrent <- dplyr::filter(allcompaniescurrent, industry == "Automotive")

ggplot(data=autocurrent, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Current
Ratio: Automotive",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#biotech
biocurrent <- dplyr::filter(allcompaniescurrent, industry == "Biotech")

ggplot(data=biocurrent, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Current
Ratio: Biotech",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#construction
constcurrent <- dplyr::filter(allcompaniescurrent, industry == "Construction")

ggplot(data=constcurrent, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Current
Ratio: Construction",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#entertainment
entercurrent <- dplyr::filter(allcompaniescurrent, industry == "Entertainment")
```

```
ggplot(data=entercurrent, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Current
Ratio: Entertainment",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#food and beverage
foodcurrent <- dplyr::filter(allcompaniescurrent, industry == "Food and Beverage")

ggplot(data=foodcurrent, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Current
Ratio: Food and Beverage",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#information technology
infocurrent <- dplyr::filter(allcompaniescurrent, industry == "Information Technology")

ggplot(data=infocurrent, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Current
Ratio: Information Technology",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#retail
retailcurrent <- dplyr::filter(allcompaniescurrent, industry == "Retail")

ggplot(data=retailcurrent, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Current
Ratio: Retail",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#sporting goods
sportcurrent <- dplyr::filter(allcompaniescurrent, industry == "Sporting Goods")

ggplot(data=sportcurrent, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Current
Ratio: Sporting Goods",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))


#quick ratios
#all companies
allcompaniesquick <- dplyr::filter(all_companies_v5, industry != "Banking",
                                   account_name == "Cash" |
                                     account_name == "Available for sale securities"|
                                     account_name == "Accounts Receivable" |
                                     account_name == "Current Liabilities")
ggplot(data=allcompaniesquick, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Quick Ratio
by industry",
    subtitle = "(Excluding Banking)", y = "% of Total Balance", x = "Year of financial reporting",
fill = "Account") +
```

```
      facet_wrap(~industry) + scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))

#aerospace
aeroquick <- dplyr::filter(allcompaniesquick, industry == "Aerospace")

ggplot(data=aeroquick, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Quick
Ratio: Aerospace",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#automotive
autoquick <- dplyr::filter(allcompaniesquick, industry == "Automotive")

ggplot(data=autoquick, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Quick
Ratio: Automotive",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#biotech
bioquick <- dplyr::filter(allcompaniesquick, industry == "Biotech")

ggplot(data=bioquick, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Quick
Ratio: Biotech",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#construction
constquick <- dplyr::filter(allcompaniesquick, industry == "Construction")

ggplot(data=constquick, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Quick
Ratio: Construction",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#entertainment
enterquick <- dplyr::filter(allcompaniesquick, industry == "Entertainment")

ggplot(data=enterquick, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Quick
Ratio: Entertainment",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#food and beverage
foodquick <- dplyr::filter(allcompaniesquick, industry == "Food and Beverage")

ggplot(data=foodquick, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Quick
Ratio: Food and Beverage",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
```

```
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#information technology
infoquick <- dplyr::filter(allcompaniesquick, industry == "Information Technology")

ggplot(data=infoquick, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Quick
Ratio: Information Technology",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#retail
retailquick <- dplyr::filter(allcompaniesquick, industry == "Retail")

ggplot(data=retailquick, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Quick
Ratio: Retail",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#sporting goods
sportquick <- dplyr::filter(allcompaniesquick, industry == "Sporting Goods")

ggplot(data=sportquick, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Quick
Ratio: Sporting Goods",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))
```

```
#Appendix E

reserveratio <- dplyr::filter(all_companies_v5, industry == "Banking", account_name == "Cash" |
                                account_name == "Customer deposits")
view(reserveratio)

ggplot(data=reserveratio, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Banking:
Reserve Ratio",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

capitalratio <- dplyr::filter(all_companies_v5, industry == "Banking", account_name ==
"Stockholder's Equity" |
                                account_name == "Loans")
view(capitalratio)

ggplot(data=capitalratio, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Banking:
Capital Ratio",
   y = "% of Total Balance", x = "Year of financial reporting", fill = "Account") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))
```

```
#Appendix F

#banking liabilities
bankliabilities <- dplyr::filter(all_companies_v5, industry == "Banking", account_type ==
"Liability",
                                 account_name != "Total Liabilities")

view(bankliabilities)

ggplot(data=bankliabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition: Banking",
    y = "% of Total Liabilities", x = "Year of financial reporting", fill = "Liability account") +
facet_wrap(~company_name) +
     scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#all other companies liabilities
liabilities <- dplyr::filter(all_companies_v5, industry != "Banking", account_name == "Current
Liabilities" |
                               account_name == "Non-current Liabilities")
view(liabilities)

ggplot(data=liabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition by industry",
  subtitle = "(Excluding Banking)",  y = "% of Total Liabilities", x = "Year of financial
reporting", fill = "Liability account") +
  facet_wrap(~industry) +  scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))

#aerospace
aeroliabilities <- dplyr::filter(liabilities, industry == "Aerospace")

ggplot(data=aeroliabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition: Aerospace",
   y = "% of Total Liabilities", x = "Year of financial reporting", fill = "Liability type") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#automotive
autoliabilities <- dplyr::filter(liabilities, industry == "Automotive")

ggplot(data=autoliabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition: Automotive",
   y = "% of Total Liabilities", x = "Year of financial reporting", fill = "Liability type") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#biotech
bioliabilities <- dplyr::filter(liabilities, industry == "Biotech")

ggplot(data=bioliabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition: Biotech",
   y = "% of Total Liabilities", x = "Year of financial reporting", fill = "Liability type") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))
```

```
#construction
constliabilities <- dplyr::filter(liabilities, industry == "Construction")

ggplot(data=constliabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition: Construction",
   y = "% of Total Liabilities", x = "Year of financial reporting", fill = "Liability type") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#entertainment
enterliabilities <- dplyr::filter(liabilities, industry == "Entertainment")

ggplot(data=enterliabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition: Entertainment",
   y = "% of Total Liabilities", x = "Year of financial reporting", fill = "Liability type") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#food and beverage
foodliabilities <- dplyr::filter(liabilities, industry == "Food and Beverage")

ggplot(data=foodliabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition: Food and Beverage",
   y = "% of Total Liabilities", x = "Year of financial reporting", fill = "Liability type") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#information technology
infoliabilities <- dplyr::filter(liabilities, industry == "Information Technology")

ggplot(data=infoliabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition: Information Technology",
   y = "% of Total Liabilities", x = "Year of financial reporting", fill = "Liability type") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#retail
retailliabilities <- dplyr::filter(liabilities, industry == "Retail")

ggplot(data=retailliabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition: Retail",
   y = "% of Total Liabilities", x = "Year of financial reporting", fill = "Liability type") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
seq(2008,2020,by=2))

#sporting goods
sportliabilities <- dplyr::filter(liabilities, industry == "Sporting Goods")

ggplot(data=sportliabilities, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Liability
composition: Sporting Goods",
   y = "% of Total Liabilities", x = "Year of financial reporting", fill = "Liability type") +
facet_wrap(~company_name) +
    scale_y_continuous(labels = scales::percent) + scale_x_continuous(breaks =
```

```
seq(2008,2020,by=2))
```

```
#Appendix G

#all companies
profitmargin <- dplyr::filter(all_companies_v5, account_name == "Net Income (Loss)" |
                                account_name == "COS and Expenses")
profitmarginnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, account_name == "Net Income
(Loss)" |
                                account_name == "COS and Expenses")
view(profitmargin)
view(profitmarginnoneg)

ggplot(data=profitmarginnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin",
    subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
  facet_wrap(~industry) + scale_y_continuous(labels = scales::percent)+ scale_x_continuous(breaks
= seq(2008,2020,by=2))

ggplot(data=profitmargin, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", show.legend = c(size=FALSE)) + labs(title = "Net Profit Margin",
  subtitle = "(as a % of Revenue)", y = "% compared to Revenue", x = "Year of financial
reporting", fill = "Account") +
  facet_wrap(~industry) +  scale_y_continuous(labels = scales::dollar_format(scale = 1e-6))

#aerospace
aeropmnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, industry == "Aerospace",
                             account_name == "Net Income (Loss)" | account_name == "COS and
Expenses")
ggplot(data=aeropmnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin: Aerospace",
  subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
  facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))

#automotive
autopmnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, industry == "Automotive",
                             account_name == "Net Income (Loss)" | account_name == "COS and
Expenses")
ggplot(data=autopmnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin: Automotive",
   subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
     facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))

#banking
bankpmnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, industry == "Banking",
                             account_name == "Net Income (Loss)" | account_name == "COS and
Expenses")
ggplot(data=bankpmnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin: Banking",
    subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
      facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#biotech
biopmnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, industry == "Biotech",
                             account_name == "Net Income (Loss)" | account_name == "COS and
```

```
Expenses")
ggplot(data=biopmnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin: Biotech",
   subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
     facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent)+
scale_x_continuous(breaks = seq(2008,2020,by=2))

#construction
constpmnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, industry == "Construction",
                            account_name == "Net Income (Loss)" | account_name == "COS and
Expenses")
ggplot(data=constpmnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin: Construction",
   subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))

#entertainment
entpmnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, industry == "Entertainment",
                              account_name == "Net Income (Loss)" | account_name == "COS and
Expenses")
ggplot(data=entpmnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin: Entertainment",
   subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))

#foodandbeverage
foodpmnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, industry == "Food and Beverage",
                            account_name == "Net Income (Loss)" | account_name == "COS and
Expenses")
ggplot(data=foodpmnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin: Food and Beverage",
   subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))

#informationtechnology
infopmnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, industry == "Information
Technology",
                             account_name == "Net Income (Loss)" | account_name == "COS and
Expenses")
ggplot(data=infopmnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin: Information Technology",
   subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))

#retail
retailpmnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, industry == "Retail",
                             account_name == "Net Income (Loss)" | account_name == "COS and
Expenses")
ggplot(data=retailpmnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin: Retail",
```

```
   subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))

#sport
sportpmnoneg <- dplyr::filter(all_companies_v5, usd_in_000s > 0, industry == "Sporting Goods",
                               account_name == "Net Income (Loss)" | account_name == "COS and
Expenses")
ggplot(data=sportpmnoneg, aes(fill = account_name, y=usd_in_000s, x = fiscal_year)) +
  geom_bar(stat="sum", position = "fill", show.legend = c(size=FALSE)) + labs(title = "Net Profit
Margin: Sporting Goods",
   subtitle = "(Expenses and Net Income as a % of Revenue)", y = "% of Revenue", x = "Year of
financial reporting", fill = "Financial Line Item") +
    facet_wrap(~company_name) + scale_y_continuous(labels = scales::percent) +
scale_x_continuous(breaks = seq(2008,2020,by=2))
```