

Analysis - Franklin Glance

How did the rise in technology usage affect historically successful businesses/industries?

```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(forcats)
library(lubridate)
library(sparklyr)
library(ggthemes)
library(gridExtra)
library(readxl)
library(stringr)
library(corr)
library(here)
library(corrgram)
library(readr)
library(ggpubr)
library(data.table)
library(modelr)
setwd("/Users/franklinglance/Library/CloudStorage/OneDrive-UniversityofVirginia/UVA/Spring
2022/SYS 2202/Final Project/Fortune_500_Analysis")
```

Importing Data

```

fortune1955_2021 <- read_csv("data/fortune500_1955_2021.csv") # rank + revenue data for f
ortune 500 companies
percent_using_internet <- read_excel("data/percent_of_population_using_internet.xlsx")

computing_efficiency <- read_csv("data/computing-efficiency.csv")
# data on various business indicators (see metadata)
data_business_indicators <- read_csv("data/data_business_indicators/aee336d2-021b-4021-912
c-3de7fd9d2729_Data.csv")

# data on supercomputer power
supercomputer_power_flops <- read_csv("data/supercomputer-power-flops.csv")

# data on phone subscriptions
landline_cellular_data <- read_excel("data/landline_cellular_data.xlsx")

# company info
company_info <- read_csv("data/Companies.csv")

# massive company dataset (7 million companies)
company_infov2 <- read_csv("data/companies_sorted.csv")

# excel sheet I created for breaking down industries into stock market sectors
industries_to_sector <- read_excel("data/industries_to_sector.xlsx")

# moores law
moores_law <- read_csv("data/transistors-per-microprocessor.csv")

# historical data on computer memory cost
computer_storage_cost <- read_csv("data/historical-cost-of-computer-memory-and-storage.cs
v")

```

Dataset Information

Dataset	Source	Description	Usage
fortune_1955_2021	https://www.kaggle.com/darinshawley/fortune-500-companies-19552021 (https://www.kaggle.com/darinshawley/fortune-500-companies-19552021)	Since 1955, Fortune magazine has published the top 500 companies in the	
US based on annual revenue. This list shows all 500 for each year. Variables: Year, Company Name, Revenue, Rank	Primary dataset, used for time series information on Fortune 500 rank + revenue.		

Dataset	Source	Description	Usage
company_infov2	https://www.kaggle.com/datasets/peopledatalabssf/free-7-million-company-dataset (https://www.kaggle.com/datasets/peopledatalabssf/free-7-million-company-dataset)	Dataset of 7+ Million Companies	Used to provide detailed information on every company in fortune_1955_2021
industries_to_sector	Self Created	Conversion Table	Used to recode industry column into stock market Sectors
Technology Adoption	https://ourworldindata.org/grapher/technology-adoption-by-households-in-the-united-states (https://ourworldindata.org/grapher/technology-adoption-by-households-in-the-united-states)	This dataset details the rates of diffusion and adoption of a range of technologies in the United States, measured as the percentage of US households with access or adoption over time.	Evaluating effect of technology adoption on businesses
Supercomputer Power	https://ourworldindata.org/grapher/supercomputer-power-flops (https://ourworldindata.org/grapher/supercomputer-power-flops)	Number of floating-point operations carried out per second by the largest supercomputer in any given year.	Evaluating effects of computational power on businesses
Computing Efficiency	https://ourworldindata.org/grapher/computing-efficiency (https://ourworldindata.org/grapher/computing-efficiency)	Computer processing efficiency, measured as the number of watts needed per million instructions per second (Watts)	Evaluating effects of computational efficiency on businesses

Dataset	Source	Description	Usage
Internet Usage	https://data.worldbank.org/indicator/IT.NET.USER.ZS (https://data.worldbank.org/indicator/IT.NET.USER.ZS)	Percent of US population using the internet by year	Evaluating effects of increasing internet usage on businesses
Phone Usage	https://data.worldbank.org/indicator/IT.CEL.SETS (https://data.worldbank.org/indicator/IT.CEL.SETS)	Data on Cellular and Land line subscriptions in the US	Evaluating effects of changing phone usage on businesses
Moore's Law	https://ourworldindata.org/technological-change (https://ourworldindata.org/technological-change)	Real world data on Moore's law, which states the number of transistors on a microchip doubles about every two years, though the cost of computers is halved.	Evaluating effects of transistor size (indicator of technological progress) on businesses
Computing Cost	https://ourworldindata.org/technological-change (https://ourworldindata.org/technological-change)	Historical cost of computer memory and storage, measured in US dollars per megabyte.	Evaluating effects of computational cost on businesses
fortune_joined	Self Created	N/A	Primary Dataset for the analysis. Contains all info on Fortune 500 Rank, Revenue, etc... Combination of every dataset in this table.

Analysis Overview

1. Prepare Dataset Representing Fortune 500 Companies
2. First round analysis on Fortune 500 dataset.
3. Clarify Assumptions/Goals
4. Quantify Success
5. Analyze relationship between technological factors and success

6. Summarize Results

Dataset Summary

Fortune 500 Data

Technological Factors Data

Part 1: Preparing Fortune 500 Dataset

Prepare Fortune 1000 and Fortune1995_2021 for joining + join by company. If missing values were present in the data, I would have removed them at this point. Additionally, the data was scanned for outliers, and there are none.

```
# add "Company" column in time scale data set on fortune 500.
fortune1955_2021 <- fortune1955_2021 %>%
  mutate(Company = Name)
# create company info dataset, holds information on 7 million companies, used to fill out
# missing information on time-scale data
company_infov2 <- company_infov2 %>%
  rename(Company = name)
```

```
# make names lowercase
lowercase_companies_1955_2021 <- fortune1955_2021 %>%
  mutate(Company = tolower(Company))
fortune_joined_v2 <- lowercase_companies_1955_2021 %>%
  left_join(company_infov2, by = "Company")
# fixing data types
#fortune_joined_v2 <- fortune_joined_v2 %>%
#  mutate(Revenue = as.numeric(Revenue))
```

Tidying Fortune Joined Dataset

Change revenue to be in billions of dollars

```
# changing revenue to numeric type, now in billions of dollars
fortune_joined_v2 <- fortune_joined_v2 %>%
  transform(Revenue = as.double(Revenue)) %>%
  mutate(Revenue = Revenue/1000)

# na values are from 2015. No free public data on fortune 500 in 2015. It is okay to be missing one year for the analysis.
fortune_joined_v2 <- fortune_joined_v2 %>%
  filter(!is.na(Revenue))
```

Part 2: First Round Analysis on Fortune 500 Dataset

What Constitutes Success?

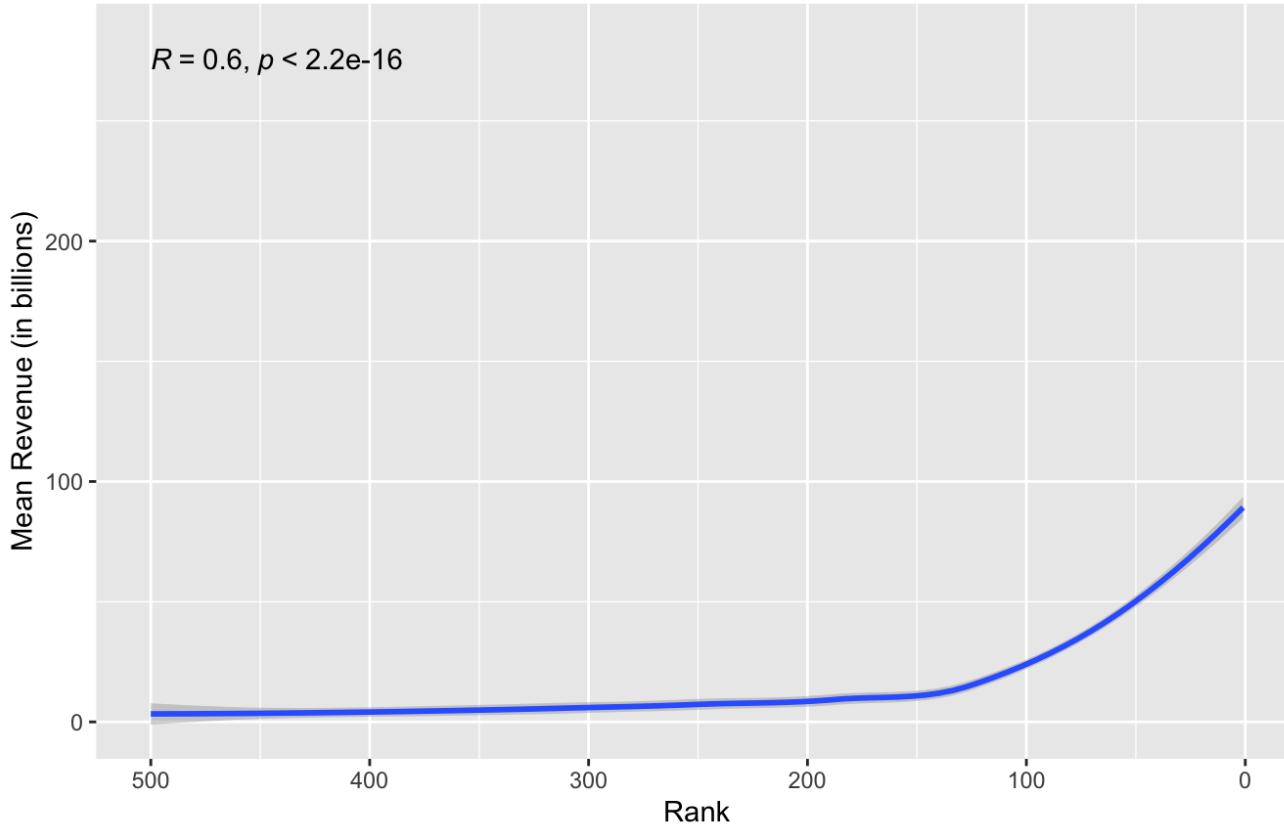
Can rank and Revenue be used synonymously when talking about success?

It appears that there exists a strong correlation between rank and revenue, so we can use either rank or revenue to quantify success. This will be useful in the future as it is easier to bin a ranking range rather than having to make buckets for all revenues. The correlation is significantly stronger when fit with a log curve, which suggests that while revenue and rank are correlated, as rank gets lower, the growth of revenue becomes much faster (even exponential). This means that a high ranking company is likely exponentially more successful (in terms of revenue) than a lower ranking company on the fortune 500.

```
fortune_joined_v2 %>%
  filter(Year > 1985) %>%
  group_by(Rank) %>%
  summarise(mean_revenue_by_rank = mean(Revenue, na.rm = TRUE)) %>%
  ggplot(aes(Rank, mean_revenue_by_rank)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Mean Revenue by Rank, Fortune 500 from 1985 to 2021",
       subtitle = "Displays correlation between rank and revenue") +
  ylab("Mean Revenue (in billions)") +
  scale_x_continuous(trans = "reverse")
```

Mean Revenue by Rank, Fortune 500 from 1985 to 2021

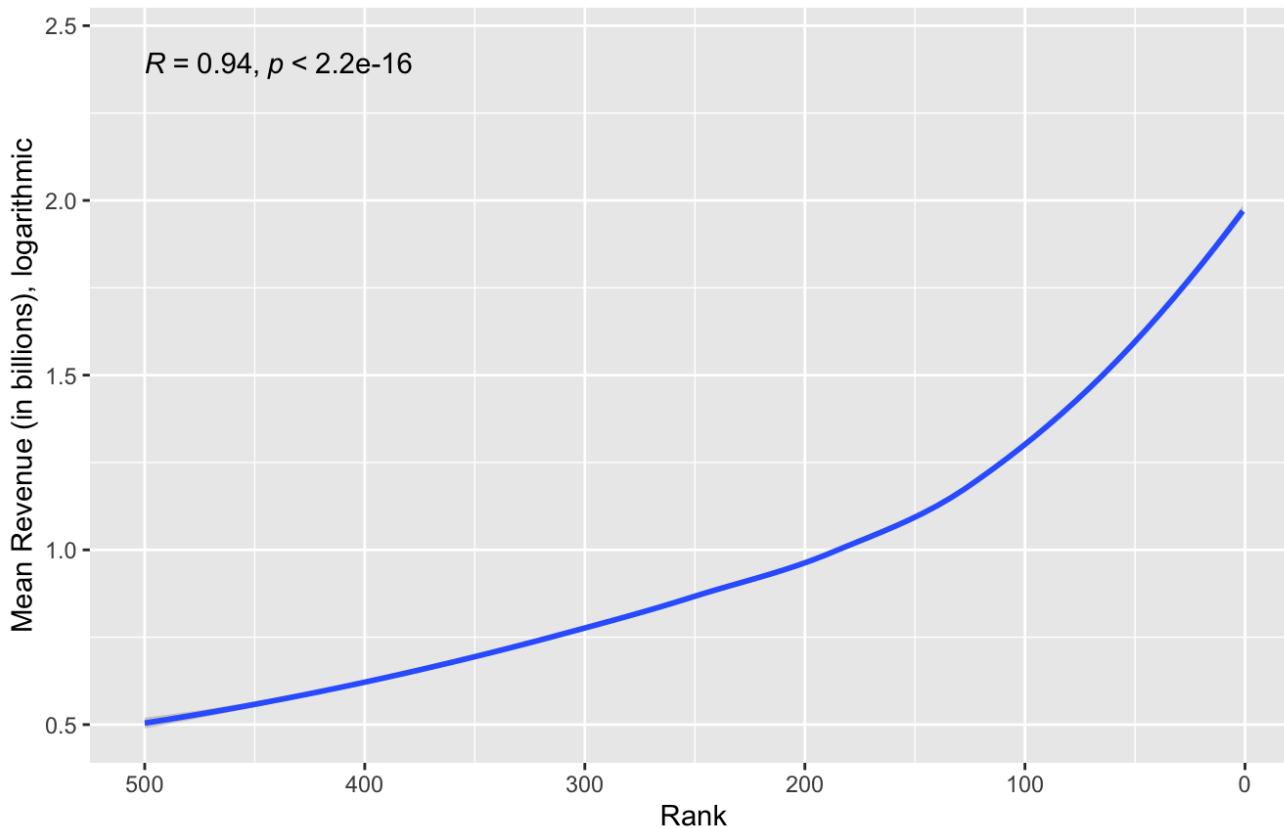
Displays correlation between rank and revenue



```
# with log curve
fortune_joined_v2 %>%
  filter(Year > 1985) %>%
  group_by(Rank) %>%
  summarise(mean_revenue_by_rank = mean(Revenue, na.rm = TRUE)) %>%
  ggplot(aes(Rank, log10(mean_revenue_by_rank))) +
  geom_smooth() +
  labs(title = "Mean Revenue by Rank, Fortune 500 from 1985 to 2021, Log curve",
       subtitle = "Displays correlation between rank and revenue") +
  ylab("Mean Revenue (in billions), logarithmic") +
  stat_cor() +
  scale_x_continuous(trans = "reverse")
```

Mean Revenue by Rank, Fortune 500 from 1985 to 2021, Log curve

Displays correlation between rank and revenue

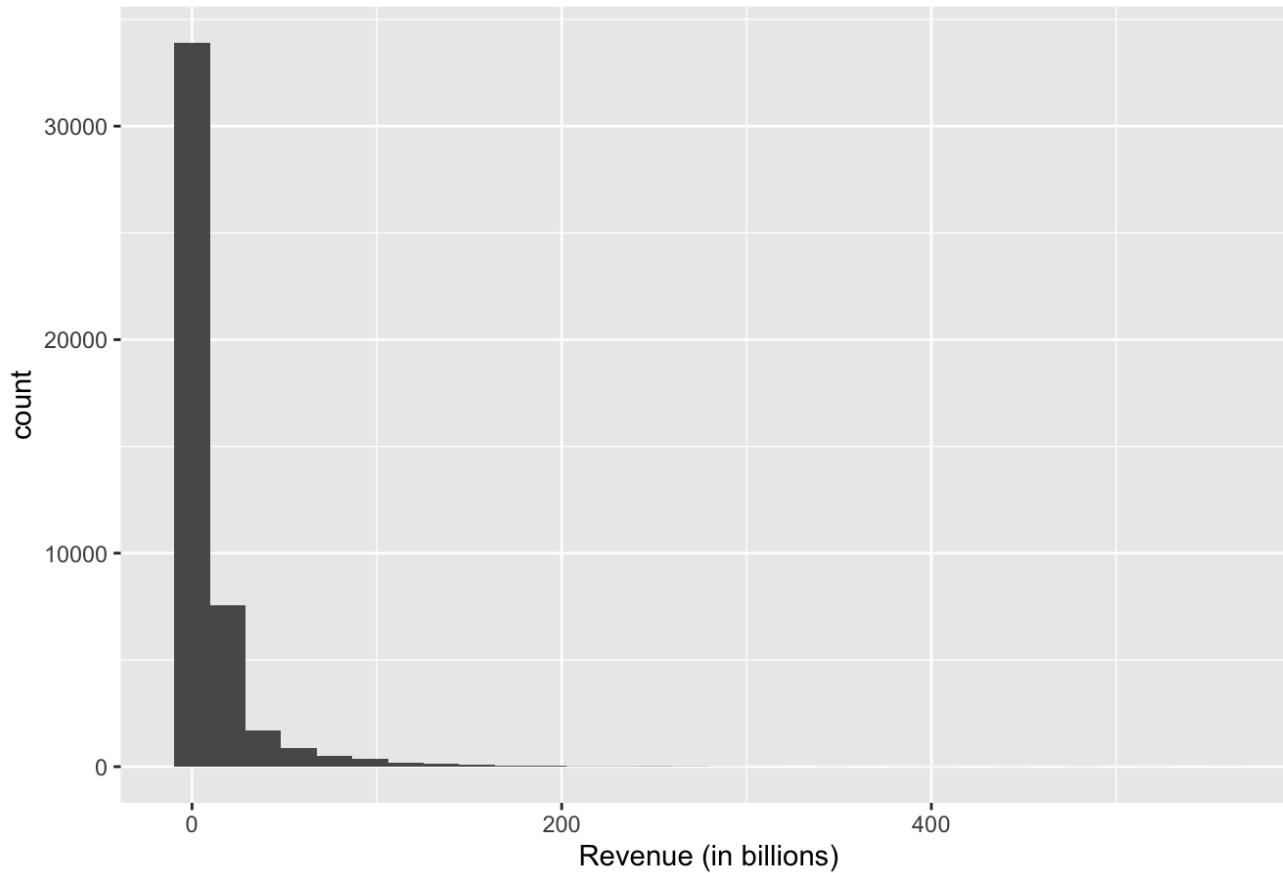


Visualization of Covariance within Revenue + Checking for Outliers and Unexpected values

This histogram would suggest that outliers exist (which is the case), however, further analysis proves that these are not outliers (they are simply extremely successful companies). The distribution of revenue is clearly extremely condensed around zero. There are no missing values (aside from the year 2015, which I know is missing).

```
fortune_joined_v2 %>%
  ggplot(aes(Revenue)) +
  geom_histogram() +
  labs(title = "Distribution of Revenue within Fortune 500 from 1985 to 2021")+
  xlab("Revenue (in billions)")
```

Distribution of Revenue within Fortune 500 from 1985 to 2021



```
#verifying not outliers, these companies make sense to have revenue values over 400 billion.
n.
```

```
fortune_joined_v2 %>%
  filter(Revenue == 0) %>%
  summarise(unique(Company))
```

```
## [1] unique(Company)
## <0 rows> (or 0-length row.names)
```

```
# checking for missing/strange values all ok
summary(fortune_joined_v2$Revenue)
```

```
##      Min.    1st Qu.     Median      Mean    3rd Qu.      Max.
## 0.0497    0.6024    3.2870   10.9581   9.8660 559.1510
```

Recoding Fortune 500 Data

Initially, I thought that I could group by industry, however there are clearly too many categories (126 to be exact).

```
fortune_joined_v2 %>%
  group_by(industry) %>%
  summarise(unique(industry))
```

```
## # A tibble: 126 × 2
##   industry      `unique(industry)`
##   <chr>        <chr>
## 1 accounting    accounting
## 2 airlines/aviation  airlines/aviation
## 3 apparel & fashion  apparel & fashion
## 4 architecture & planning  architecture & planning
## 5 arts and crafts  arts and crafts
## 6 automotive     automotive
## 7 aviation & aerospace  aviation & aerospace
## 8 banking        banking
## 9 biotechnology  biotechnology
## 10 broadcast media broadcast media
## # ... with 116 more rows
```

Since there are 126 unique industries, it is necessary to recode the Fortune 500 dataset to have a larger overarching category with which to analyze trends in the companies. It makes the most sense to use the 11 sectors of the stock market for grouping. The stocks of companies within the same industry will typically trade in the same direction, providing a good reference to the industry as a whole. Additionally, companies in the same industry are often affected by the same (or similar) factors.

Reference for Sector Breakdown: <https://time.com/nextadvisor/investing/stock-market-sectors/>
[\(https://time.com/nextadvisor/investing/stock-market-sectors/\)](https://time.com/nextadvisor/investing/stock-market-sectors/)

The Sectors Are: Energy Materials Industrials Utilities Healthcare Financials Consumer Discretionary Consumer Staples Information Technology Communication Services Real Estate

I created an excel spreadsheet mapping the current “industries” to stock market sectors.

```
# 126 unique industries
industries<- fortune_joined_v2 %>%
  summarise(unique(industry))
# export to csv
write.csv(industries, "industries.csv")
# use industries_to_sector dataframe to recode industry column
industries_to_sector <- industries_to_sector %>%
  mutate(industry = Industry)
# fortune joined now contains sector info
fortune_joined_v3 <- fortune_joined_v2 %>%
  left_join(industries_to_sector, by = "industry")
fortune_joined_v3 <- fortune_joined_v3 %>%
  filter(Sector != "NA")
# fixing data types
fortune_joined_v3 <- fortune_joined_v3 %>%
  mutate(Year = as.numeric(Year),
        Revenue = as.double(Revenue))
```

This process produces fortune_joined_v3, the primary dataset for the rest of the analysis.

Visualizing + Analyzing Fortune 500 after Sector Grouping

Which companies have been consistently on the Fortune 500?

First, we will list companies which have been on the fortune 500 for more than 20 years, after 1985. This will give us an idea of how many companies stay on the fortune 500 for an extended period of time.

Results: 317 companies have been on the Fortune 500 for at least 20 years, since 1985. This is over half of the Fortune 500.

```
head(fortune_joined_v3 %>% count(Company, sort=TRUE), n=20)
```

```
##               Company   n
## 1             ibm 462
## 2             cbs 354
## 3            harris 288
## 4            merck 264
## 5            mosaic 255
## 6  hewlett-packard 252
## 7 state farm insurance 234
## 8            fmc 220
## 9            itt 216
## 10           singer 204
## 11 international paper 198
## 12 northrop grumman 198
## 13          pepsico 198
## 14 united technologies 192
## 15           eaton 174
## 16           trw 174
## 17 baker hughes 172
## 18           amp 170
## 19          tenneco 165
## 20 american standard 159
```

```
# filtering by companies on the list for 20 years, since 1985
output <- fortune_joined_v3 %>% # hiding output for knit
  group_by(Company) %>%
  filter(Year > 1985) %>%
  filter(n() > 20) %>%
  summarise(unique(Company), n()) %>%
  arrange(desc(n()))
```

Now we will filter for companies who have been in the Fortune 500 the entire time.

Results: 168 companies have been on the fortune 500 for the entire time (1985-2021). The 168 companies who have been on the Fortune 500 would typically be large corporations, who have been around for awhile. These companies are very successful, but may be less innovative due to their constant market presence, which requires that they please shareholders. This often results in companies sticking to their tried and true methods, rather than taking the innovative leaps required to create a unicorn startup.

```
output2 <- fortune_joined_v3 %>% # hiding output for knit
  group_by(Company) %>%
  filter(Year > 1985) %>%
  filter(n() > 35) %>%
  summarise(unique(Company))
```

I will continue with selecting subsets of the Fortune 500 in the next section.

Boolean Select Columns

Next, Boolean Select columns are created within fortune_joined_v3 holding a Boolean value representing presence in several different subsets of the Fortune 500:

Subset	Description
Incumbent	Companies that have been in the Fortune 500 every year since 1985.
20+	Companies that have been in the Fortune 500 for at least 20 years since 1985.
IT	Companies within the Information Technology Sector
Startup	Companies that were founded after 2000, and are in the Fortune 500 in 2021.

```
#incumbent column
fortune_joined_v3 <- fortune_joined_v3 %>%
  group_by(Company) %>%
  filter(Year > 1985) %>%
  mutate(incumbent = ifelse(n()>35, "True", "False"))

# twenty year column
fortune_joined_v3 <- fortune_joined_v3 %>%
  group_by(Company) %>%
  filter(Year > 1985) %>%
  mutate(on_for_twenty = ifelse(n()>20, "True", "False"))

# information technology
fortune_joined_v3 <- fortune_joined_v3 %>%
  group_by(Company) %>%
  mutate(IT = ifelse(Sector == "Information Technology", "True", "False"))

# current fortune 500 companies
current <- fortune_joined_v3 %>%
  filter(Year == 2021)

# adding column holding boolean value
fortune_joined_v3 <- fortune_joined_v3 %>%
  mutate(current_fortune = ifelse(Company %in% current$Company, "True", "False"))

# adding startup
startups <- fortune_joined_v3 %>%
  group_by(Company) %>%
  filter(current_fortune == "True") %>%
  filter(year.founded > 1999)

# adding startup column
fortune_joined_v3 <- fortune_joined_v3 %>%
  mutate(startup = ifelse(Company %in% startups$Company, "True", "False"))
```

Visualizing New Subsets of the Data

Now that we have a way of narrowing in on subsets of the fortune 500 (incumbent, on_for_twenty, IT, current_fortune, startup), it is possible to visualize certain parts of data to learn more about how the Fortune 500 companies have performed over the years.

All Companies:

Average Revenue of Fortune 500 Companies from 1985 to 2021

This visualization shows the distribution of average revenue for a fortune 500 company in each sector from 1985 to 2021. It will serve as a baseline for the analysis.

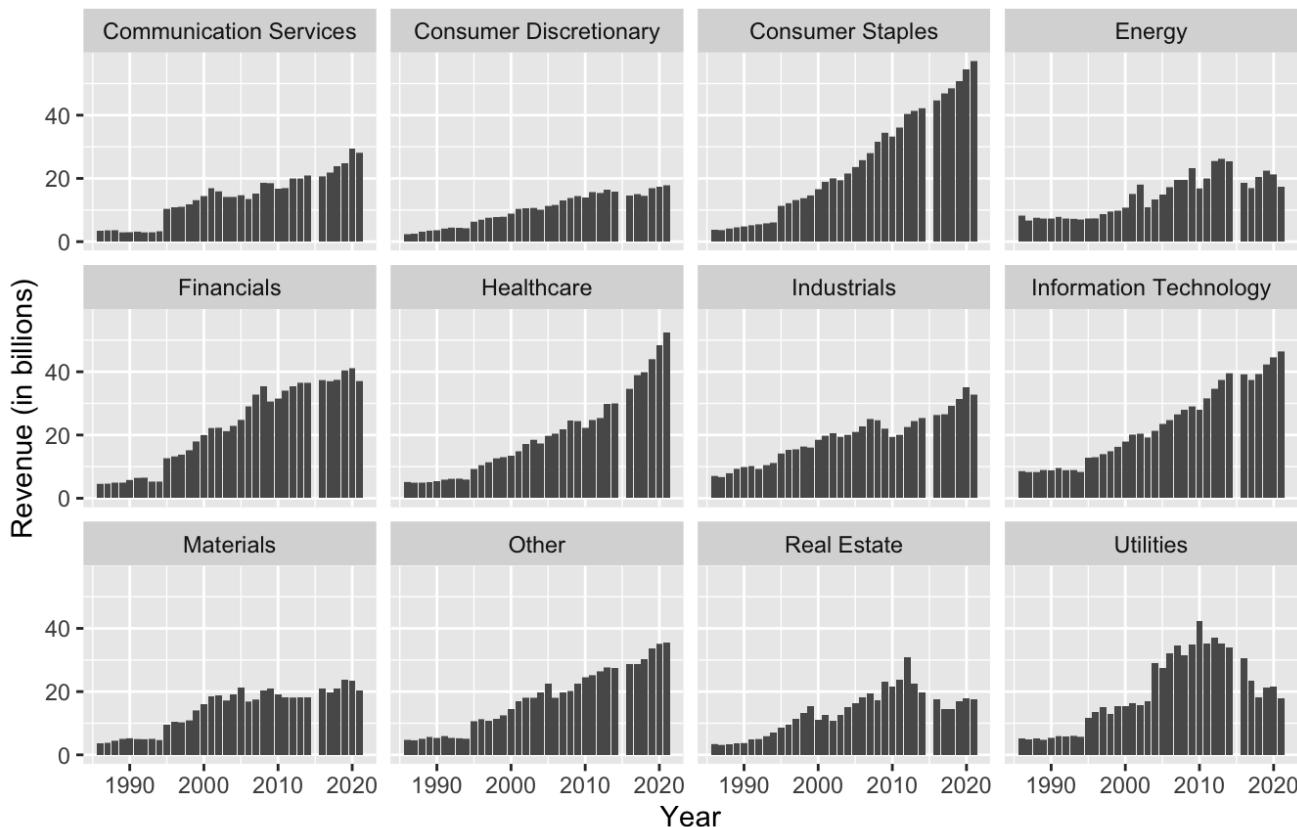
Interpretation:

- The Consumer Staples, Information Technology, and Healthcare sectors appear to have the most continuous and largest amounts of growth. This indicates that it may be more likely for a company to be successful if it is in one of these three sectors.
- Real Estate, Materials, Utilities, and Energy Sectors have not shown promising growth in recent years

```
fortune_joined_v3 %>%
  group_by(Sector, Year) %>%
  mutate(count = n()) %>%
  ggplot(aes(Year, Revenue/count)) +
  geom_col() +
  facet_wrap(~ Sector) +
  labs(title = "Average Revenue of Fortune 500 Companies from 1985 to 2021",
       subtitle = "Displays distribution of Revenue, by sector") +
  ylab("Revenue (in billions)")
```

Average Revenue of Fortune 500 Companies from 1985 to 2021

Displays distribution of Revenue, by sector



Incumbent Companies:

Average Revenue for Incumbent Fortune 500 Companies

Incumbent - Companies that have been on the Fortune 500 every year since 1985

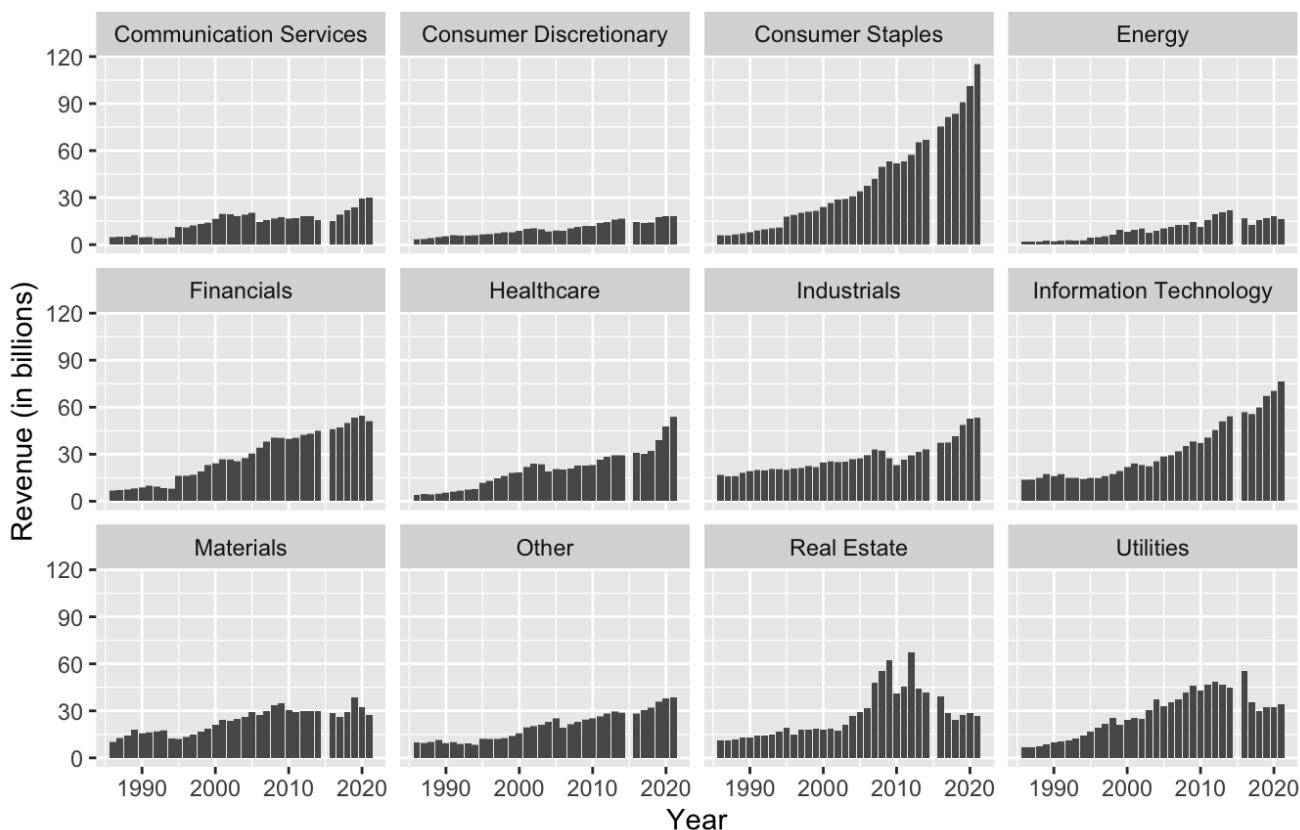
Interpretation:

- Incumbent companies have a greater average revenue than the Fortune 500 as a whole. This indicates that Incumbents are more successful on average than companies who haven't been on the list the entire time. This make sense as I would expect for the incumbents to be in the top half of the Fortune 500 in order to not be eliminated on a down year.
- Consumer Staples and Information look strong as ever
- Healthcare appears to be a more recent addition to the Fortune 500, since the revenue of Healthcare companies has grown significantly more in non-incumbent companies. This suggests that new technologies could have lead to an increase in revenue in that sector. This point will be pursued further in part tk.

```
fortune_joined_v3 %>%
  filter(incumbent == "True") %>%
  group_by(Sector, Year) %>%
  mutate(count = n()) %>%
  ggplot(aes(Year, Revenue/count)) +
  geom_col() +
  facet_wrap(~ Sector) +
  labs(title = "Average Revenue for Incumbent Fortune 500 Companies from 1985 to 2021",
       subtitle = "Displays distribution of Revenue, by sector") +
  ylab("Revenue (in billions)")
```

Average Revenue for Incumbent Fortune 500 Companies from 1985 to 2021

Displays distribution of Revenue, by sector



20 Year Companies:

Average Revenue for 20+ Fortune 500 Companies

20+ Definition - Companies that have been on the Fortune 500 for at least 20 of the 36 years since 1985.

Interpretation:

- Consumer Staples and Information Technology look strong as ever
- The financial sector appears to hold steady, almost non-decreasing growth. This suggests that the successful financial companies are likely to join the fortune 500, and stay on once they do.
- Healthcare is very strong in this subset of the Fortune 500 companies. Again, this suggests that the healthcare sector has experienced solid growth in recent years.

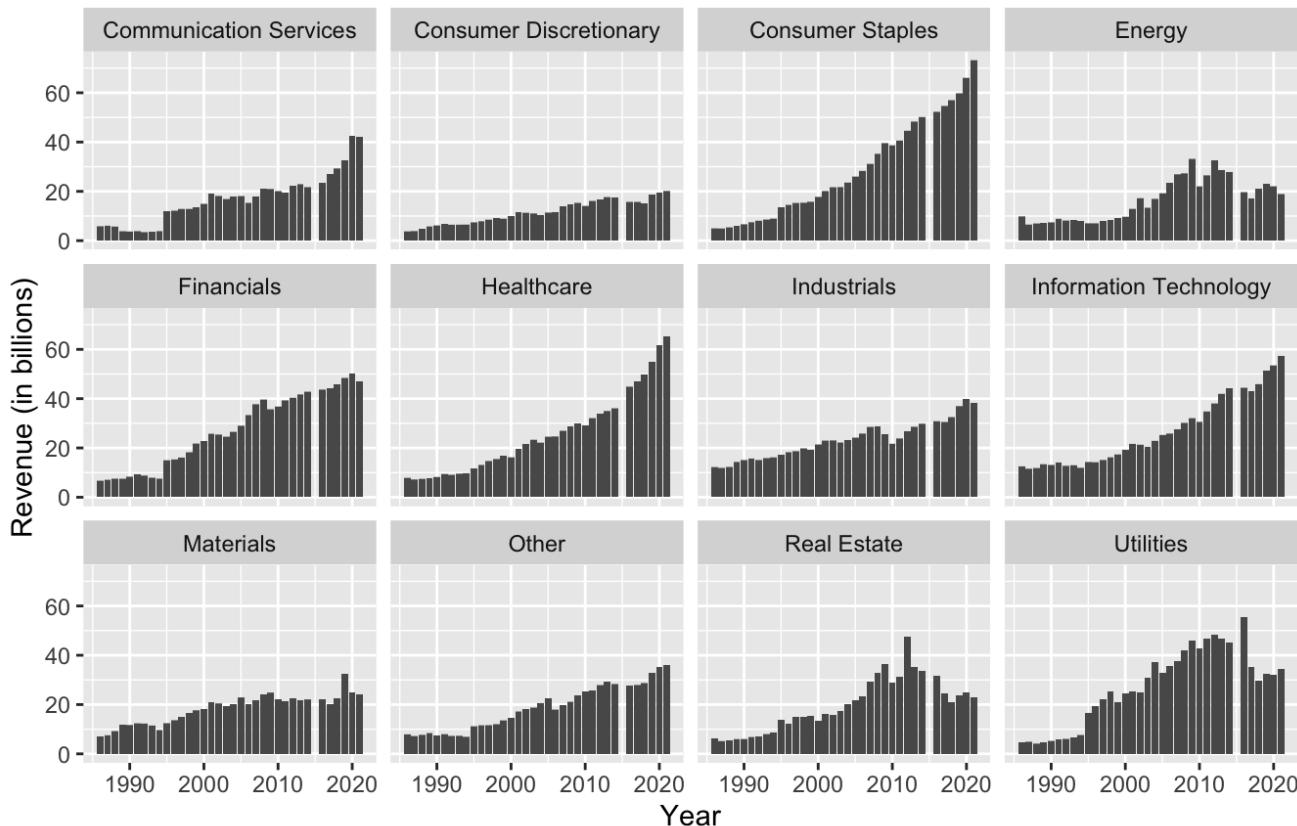
```
fortune_joined_v3
```

```
## # A tibble: 18,945 × 22
## # Groups: Company [974]
##   Year Name Revenue Rank Company ...1 domain year.founded industry
##   <dbl> <chr>    <dbl> <dbl> <chr>    <dbl> <chr>      <dbl> <chr>
## 1 2021 Walmart  559.     1 walmart  5944912 walmartcar...  1962 retail
## 2 2021 Walmart  559.     1 walmart  2637561 <NA>          NA retail
## 3 2021 Amazon   386.     2 amazon   4306268 amazon.com   1994 internet
## 4 2021 Amazon   386.     2 amazon   6535892 <NA>          NA warehou...
## 5 2021 Apple    275.     3 apple    5735407 apple.com   1976 consume...
## 6 2021 Apple    275.     3 apple    2389734 dorado.com  1998 compute...
## 7 2021 Apple    275.     3 apple    5303562 <NA>          NA consume...
## 8 2021 Apple    275.     3 apple    4507089 intelpost.... NA industr...
## 9 2021 CVS      269.     4 cvs      5614397 <NA>          NA retail
## 10 2021 CVS     269.     4 cvs      4908121 <NA>          NA retail
## # ... with 18,935 more rows, and 13 more variables: size.range <chr>,
## #   locality <chr>, country <chr>, linkedin.url <chr>,
## #   current.employee.estimate <dbl>, total.employee.estimate <dbl>,
## #   Industry <chr>, Sector <chr>, incumbent <chr>, on_for_twenty <chr>,
## #   IT <chr>, current_fortune <chr>, startup <chr>
```

```
fortune_joined_v3 %>%
  filter(on_for_twenty == "True") %>%
  group_by(Sector, Year) %>%
  mutate(count = n()) %>%
  ggplot(aes(Year, Revenue/count)) +
  geom_col() +
  facet_wrap(~ Sector) +
  labs(title = "Average Revenue for 20+ Fortune 500 Companies from 1985 to 2021",
       subtitle = "20+ -> Companies which have been on the Fortune 500 for more than 20 years") +
  ylab("Revenue (in billions)")
```

Average Revenue for 20+ Fortune 500 Companies from 1985 to 2021

20+ -> Companies which have been on the Fortune 500 for more than 20 years



IT:

Average revenue for Information Technology Fortune 500 Companies

IT - Companies in the Information Technology stock market sector. I have chosen to view IT individually because the overarching analysis is focused on analyzing the effects of Technological changes on Fortune 500 companies, so it makes sense to pay close attention to this sector in particular.

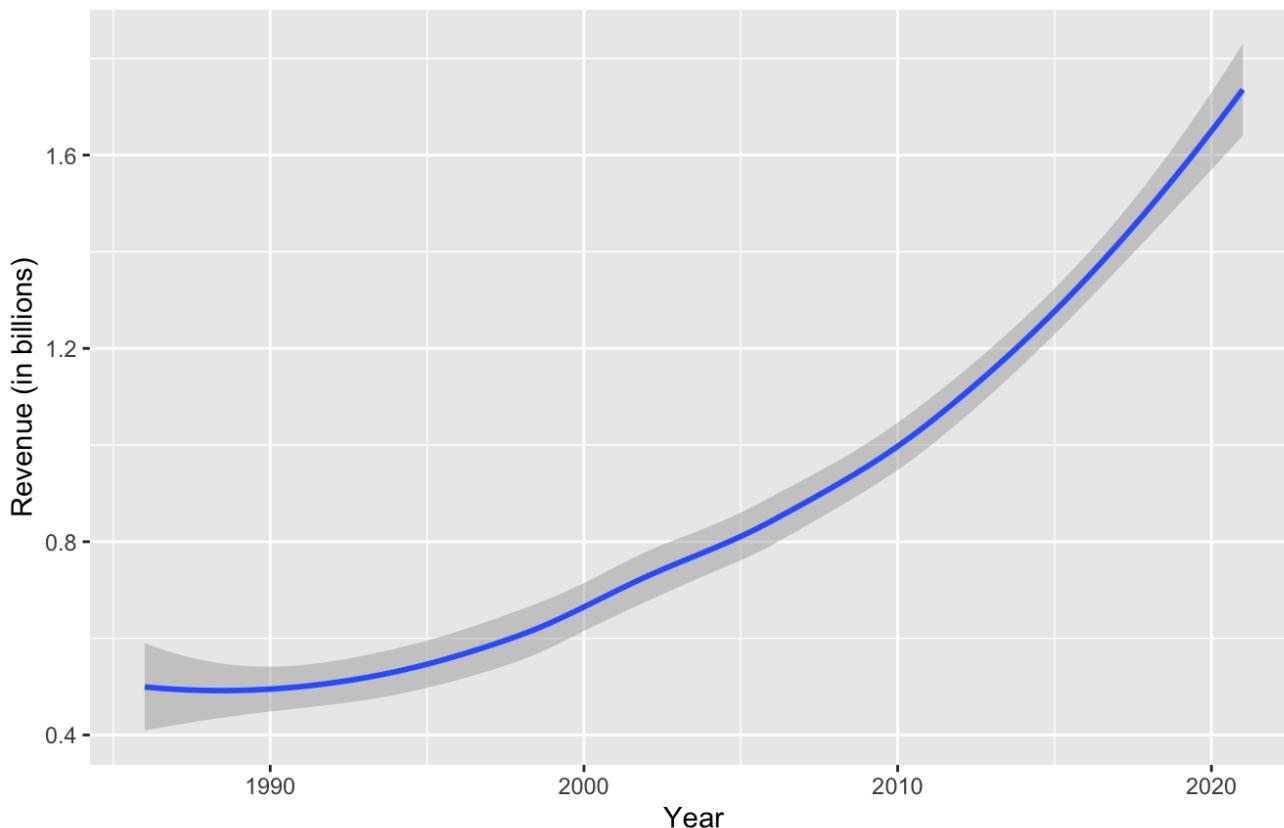
Interpretation:

- This graph provides a closer look at the Information Technology sector.
- The IT sector shows non-decreasing growth year on year.
-

```
fortune_joined_v3 %>%
  filter(IT == "True") %>%
  group_by(Sector, Year) %>%
  mutate(count = n()) %>%
  filter(Rank < 20) %>%
  ggplot(aes(Year, Revenue/count)) +
  geom_smooth() +
  labs(title = "Average Revenue for IT Fortune 500 Companies from 1985 to 2021",
       subtitle = "IT -> Information Technology Sector") +
  ylab("Revenue (in billions)")
```

Average Revenue for IT Fortune 500 Companies from 1985 to 2021

IT -> Information Technology Sector



Current Fortune 500 Companies:

Average Revenue for Fortune 500 Companies in 2021

Current Fortune 500 Companies - Companies that are currently on the Fortune 500 in 2021.

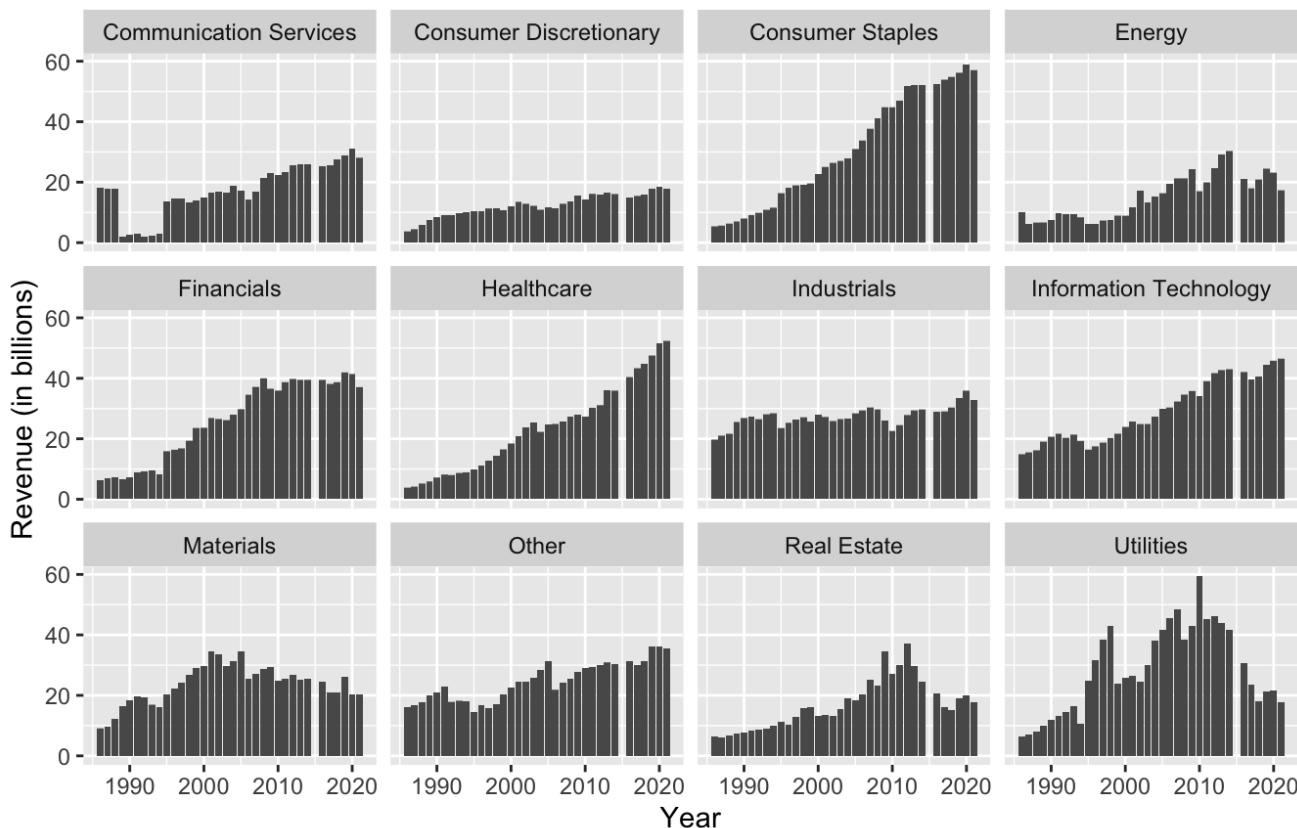
Interpretation:

- Current Fortune 500 Companies show the history of the most successful corporations in the world to date.
- Sectors like Industrials, Materials, and Utilities have struggled in recent years, while Consumer Staples, IT, Financials, and Healthcare have experienced massive growth. This is indicative of a shift in the makeup of the Fortune 500.

```
fortune_joined_v3 %>%
  filter(current_fortune == "True") %>%
  group_by(Sector, Year) %>%
  mutate(count = n()) %>%
  ggplot(aes(Year, Revenue/count)) +
  geom_col() +
  facet_wrap(~ Sector) +
  labs(title = "Average Revenue for Current Fortune 500 Companies from 1985 to 2021",
       subtitle = "Displays distribution of Revenue, by sector") +
  ylab("Revenue (in billions)")
```

Average Revenue for Current Fortune 500 Companies from 1985 to 2021

Displays distribution of Revenue, by sector



Startup Companies:

Average Revenue for Fortune 500 Startups since 2000

Startup - A company that has been founded after the year 2000, and is on the Fortune 500 in 2021.

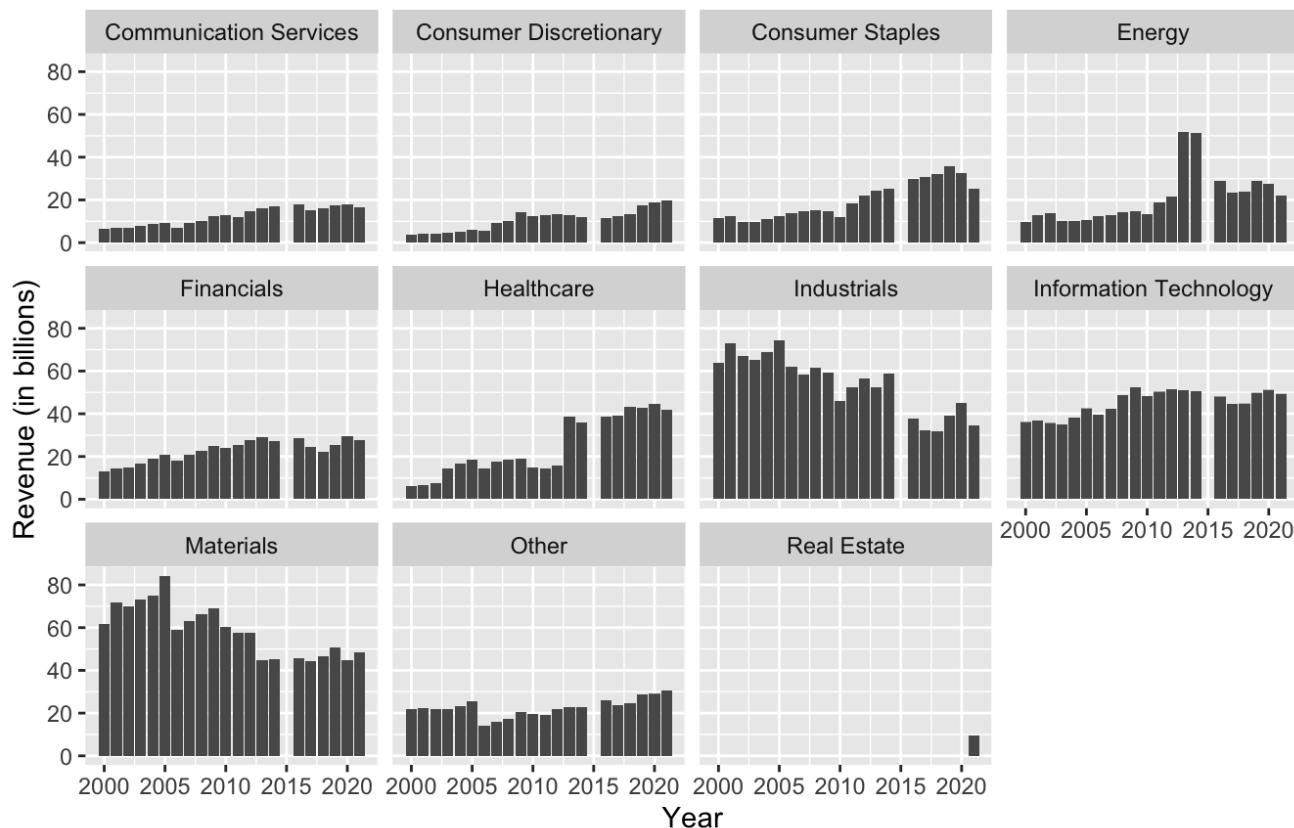
Interpretation:

- This is a subset of *Current Fortune 500 Companies*.
- It isn't surprising that Startup companies *on average* have experienced little growth on the Fortune 500 over the years, because Startups may often take much longer than 20 years to become massively successful, and a majority do fail.

```
fortune_joined_v3 %>%
  filter(startup == "True") %>%
  filter(Year > 1999) %>%
  group_by(Sector, Year) %>%
  mutate(count = n()) %>%
  ggplot(aes(Year, Revenue/count)) +
  geom_col() +
  facet_wrap(~ Sector) +
  labs(title = "Average Revenue for Fortune 500 Startups from 2000 to 2021",
       subtitle = "Displays distribution of Revenue, by sector") +
  ylab("Revenue (in billions)")
```

Average Revenue for Fortune 500 Startups from 2000 to 2021

Displays distribution of Revenue, by sector

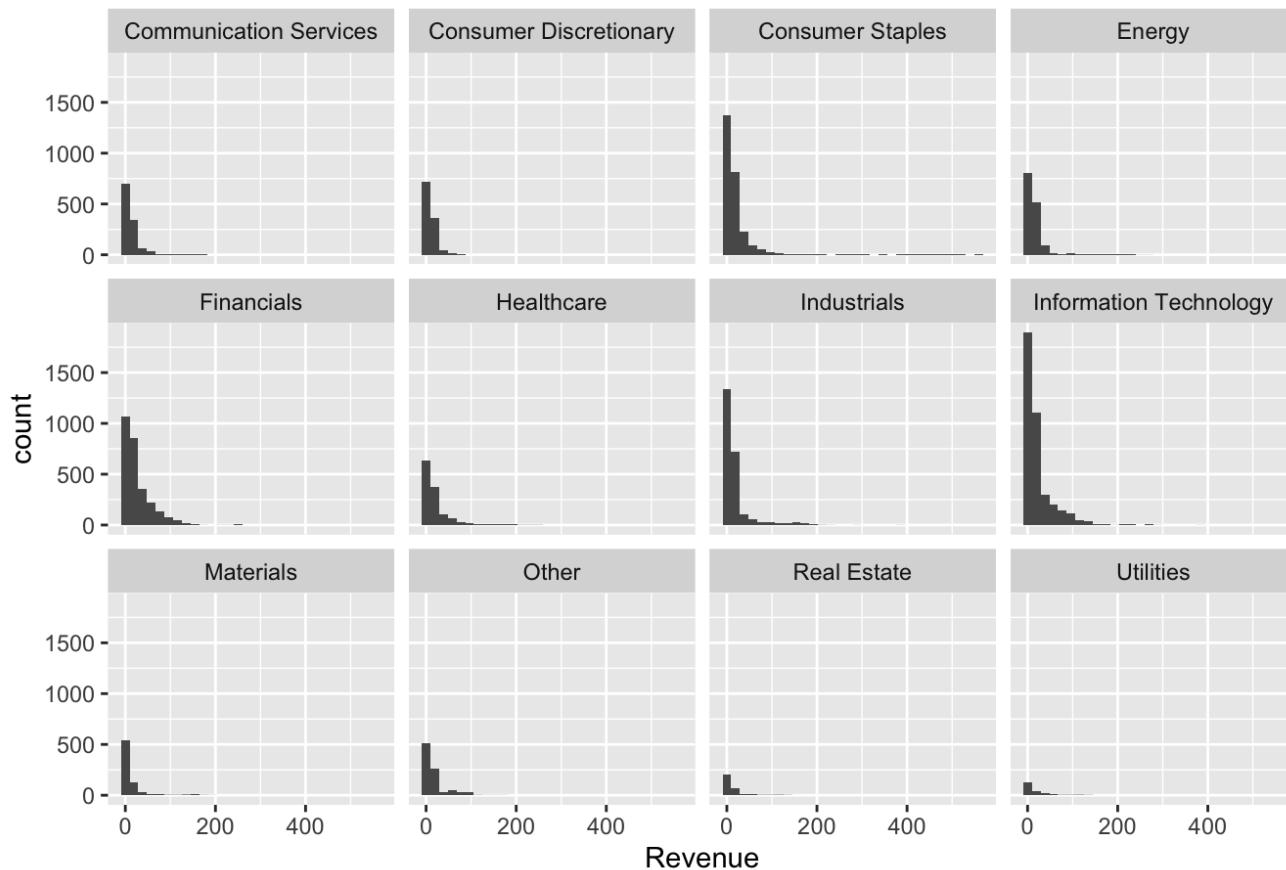


Single Variable Variation within Fortune 500 Dataset

This section served to evaluate the variation of revenue within different sectors of the Fortune 500 Dataset. This is an important step in determining the distribution of data. This visualization provides powerful insight into the relative frequency of each sector occurring on the Fortune 500, with Information Technology clearly being the most frequent.

```
fortune_joined_v3 %>%
  ggplot(aes(Revenue)) +
  geom_histogram() +
  facet_wrap(~Sector) + labs(title = "Variation within Revenue for Current Fortune 500 Companies from 1985 to 2021")
```

Variation within Revenue for Current Fortune 500 Companies from 1985 to 202



Part 3: Clarifying Assumptions/Goals

Overarching Question: **How did the rise in technology usage (social media, internet, etc..) affect historically successful businesses/industries?**

Parts 1 and 2 of this Analysis serve as the foundation of the predictive half. In the second half of this analysis, I hope to use the dataset I have just built to answer the overarching question.

The goal of this analysis is to determine the impact of technological advancements on “successful” businesses. Therefore, the second half of the analysis will cover the following steps:

- Part 4: Quantifying Success: *What subset of Fortune 500 companies are successful? unsuccessful?*
- Part 5: Technological Influences: *How has technology affected successful businesses? What does this mean for the future?*
- Part 6: Summary of Results

Part 4: Quantifying Success

The goal for this section is to narrow in on a specific selection of *successful* companies. In the Boolean select portion of the analysis, I visualized the general trends within the different subsets of the Fortune 500. This was the first step towards determining what constitutes a *successful* company. Now, I will find several companies within each subset of the Fortune 500 that model *success*.

Important Qualities:

- Non-decreasing Revenue per 5 year period (to account for select poor years)
- Currently on Fortune 500 in 2021
- Subsets:
 - Information Technology Companies
 - Healthcare Companies
 - Communication Services Companies
 - Consumer Discretionary Companies
 - Consumer Staples Companies
 - Energy Companies
 - Financial Companies
 - Healthcare Companies
 - Industrial Companies
 - Information Technology Companies
 - Materials Companies
 - Real Estate Companies
 - Each of these subsets will contain examples of both successful and unsuccessful companies

Procedure:

1. Determine specific filters for success.
2. Determine specific filters for lack of success.
3. Modify Fortune 500 Dataset to hold Boolean values representing containment in the aforementioned subsets.

Filtering for Success

Quantifying Success

Category	Qualifications
Hyper-Successful	>5 billion/year avg. revenue growth + in Fortune 500 in 2021
Moderately-Successful	>1 && <5 billion/year avg. revenue growth + in Fortune 500 in 2021
Low-Success	<1 billion/year avg. revenue growth + in Fortune 500 in 2021
Unsuccessful*	<.5 billion/year avg. revenue growth + not in Fortune 500 in 2021

```

# add max and min revenue value for each company
fortune_joined_v3 <- fortune_joined_v3 %>%
  group_by(Company) %>%
  filter(Year > 1985) %>%
  mutate(max_revenue = max(Revenue, na.rm = TRUE),
        min_revenue = min(Revenue, na.rm = TRUE),
        revenue_diff = max_revenue - min_revenue,
        revenue_delta = revenue_diff/(max(Year, na.rm = TRUE) - min(Year, na.rm = TRUE)),
        years_on_fortune = max(Year, na.rm = TRUE) - min(Year, na.rm = TRUE))

# max_revenue -> maximum revenue during timeframe on fortune 500
# min_revenue -> minimum revenue during timeframe on fortune 500
# revenue diff -> change in revenue while on fortune 500
# revenue delta -> change in revenue (in billions/year) while on fortune 500

# create hyper-growth company dataset (<5 billion/year revenue growth, on average)
hyper_growthdf <- fortune_joined_v3 %>%
  filter(current_fortune == "True") %>% # company must be on Fortune 500 in 2021
  group_by(Company) %>%
  filter(years_on_fortune > 10) %>% # need at least 10 years of data for confidence, down
  to 262 companies
  filter(!Sector %in% c("Other", "NA")) %>%
  filter(revenue_delta > 5)

# create moderate-growth company dataset ( >1 && <5 billion/year revenue growth, on average)
moderate_growthdf <- fortune_joined_v3 %>%
  filter(current_fortune == "True") %>% # company must be on Fortune 500 in 2021
  group_by(Company) %>%
  filter(years_on_fortune > 10) %>% # need at least 10 years of data for confidence
  filter(!Sector %in% c("Other", "NA")) %>%
  filter(revenue_delta > 1 && revenue_delta < 5)

# create low-growth company dataset (<1 billion/year revenue growth, on average)
low_growthdf <- fortune_joined_v3 %>%
  filter(current_fortune == "True") %>% # company must be on Fortune 500 in
  group_by(Company) %>%
  filter(years_on_fortune > 10) %>% # need at least 10 years of data for confidence
  filter(!Sector %in% c("Other", "NA")) %>%
  filter(revenue_delta < 1 )

# add boolean value for hyper-growth to fortune 500 dataset
fortune_joined_v3 <- fortune_joined_v3 %>%
  mutate(hyper_growth = ifelse(Company %in% hyper_growthdf$Company, "True", "False"))

# add boolean value for moderate-growth to fortune 500 dataset
fortune_joined_v3 <- fortune_joined_v3 %>%
  mutate(moderate_growth = ifelse(Company %in% moderate_growthdf$Company, "True", "False"))

# add boolean value for low-growth to fortune 500 dataset

```

```
fortune_joined_v3 <- fortune_joined_v3 %>%
  mutate(low_growth = ifelse(Company %in% low_growthdf$Company, "True", "False"))
```

Summary:

- Added columns containing info on low, moderate, and high growth. These will be the Boolean conditions for “success”.

Filtering for Unsuccessful Companies

```
# create unsuccessful company dataset (<.5 billion/year revenue growth, on average) + not
currently on Fortune 500 + between 10 and 20 years on list
unsuccessful_df <- fortune_joined_v3 %>%
  filter(current_fortune == "False") %>% # company must not be on Fortune 500 in 2021
  group_by(Company) %>%
  filter(years_on_fortune > 10 && years_on_fortune < 20) %>% # need at least 10 years of d
ata for confidence, but less than 20 years,
  filter(!Sector %in% c("Other", "NA")) %>%
  filter(revenue_delta < .5)

# add boolean value for unsuccessful to fortune 500 dataset
fortune_joined_v3 <- fortune_joined_v3 %>%
  mutate(unsuccessful = ifelse(Company %in% unsuccessful_df$Company, "True", "False"))
```

Visualizing Success

We now have Boolean values that can be used to filter down to various successful/unsuccessful companies.

Hyper Growth Companies

These companies are the most successful of the bunch. Because Hyper-Growth will remain the main focus of the rest of the analysis, it necessary to identify the makeup of the companies. The following plots serve that purpose.

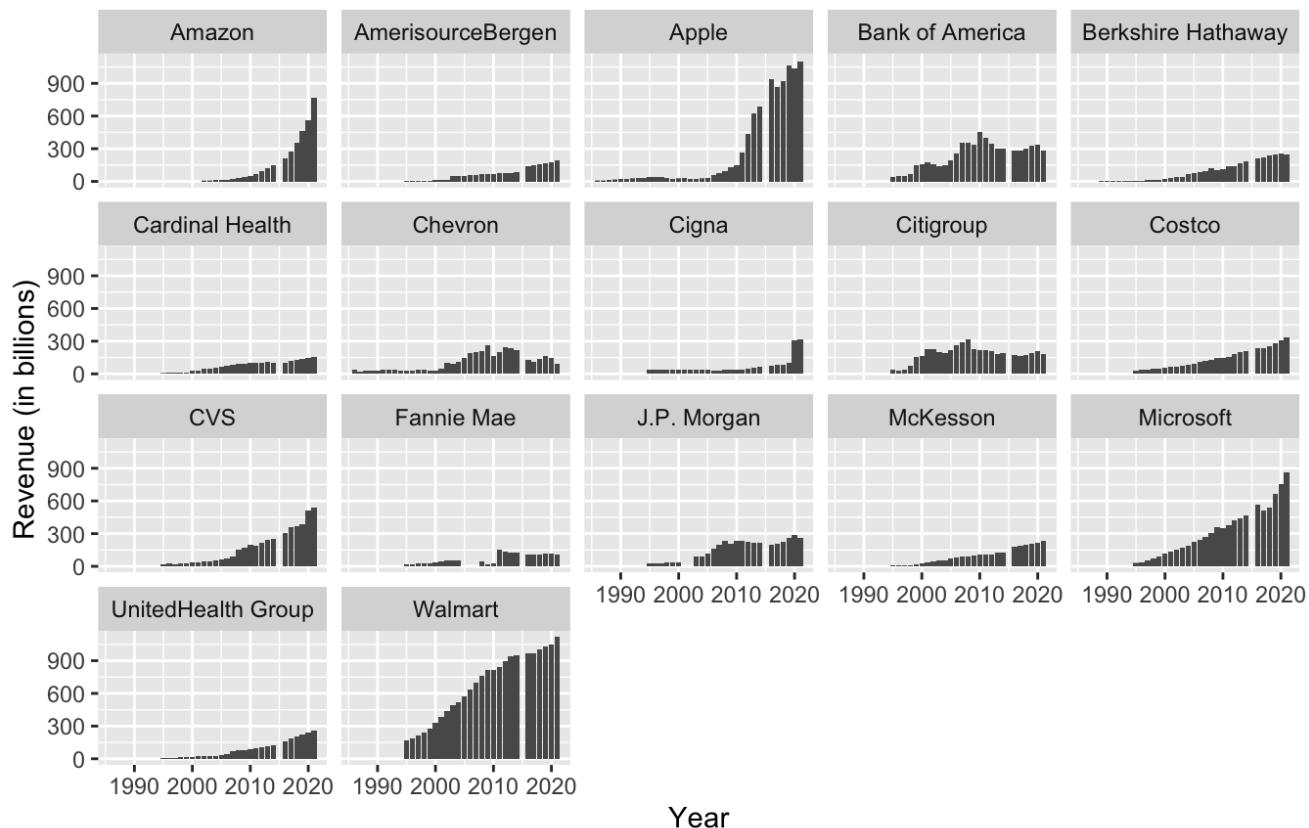
Results:

- These companies exhibit extraordinary growth, and are very recognizable names.
- They fall into the Consumer Staples, Energy, Financials, Healthcare and Information Technology Sectors.
- Information Technology is the only sector to exhibit exponential growth. This suggests that Information Technology companies have the most potential for rapid expansion.

```
# plotting all names
fortune_joined_v3 %>%
  filter(hyper_growth == "True") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col() +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Hyper-Growth Companies from 1985 to 2021",
       subtitle = "Displays distribution of Revenue, by company") +
  ylab("Revenue (in billions)")
```

Revenue for Hyper-Growth Companies from 1985 to 2021

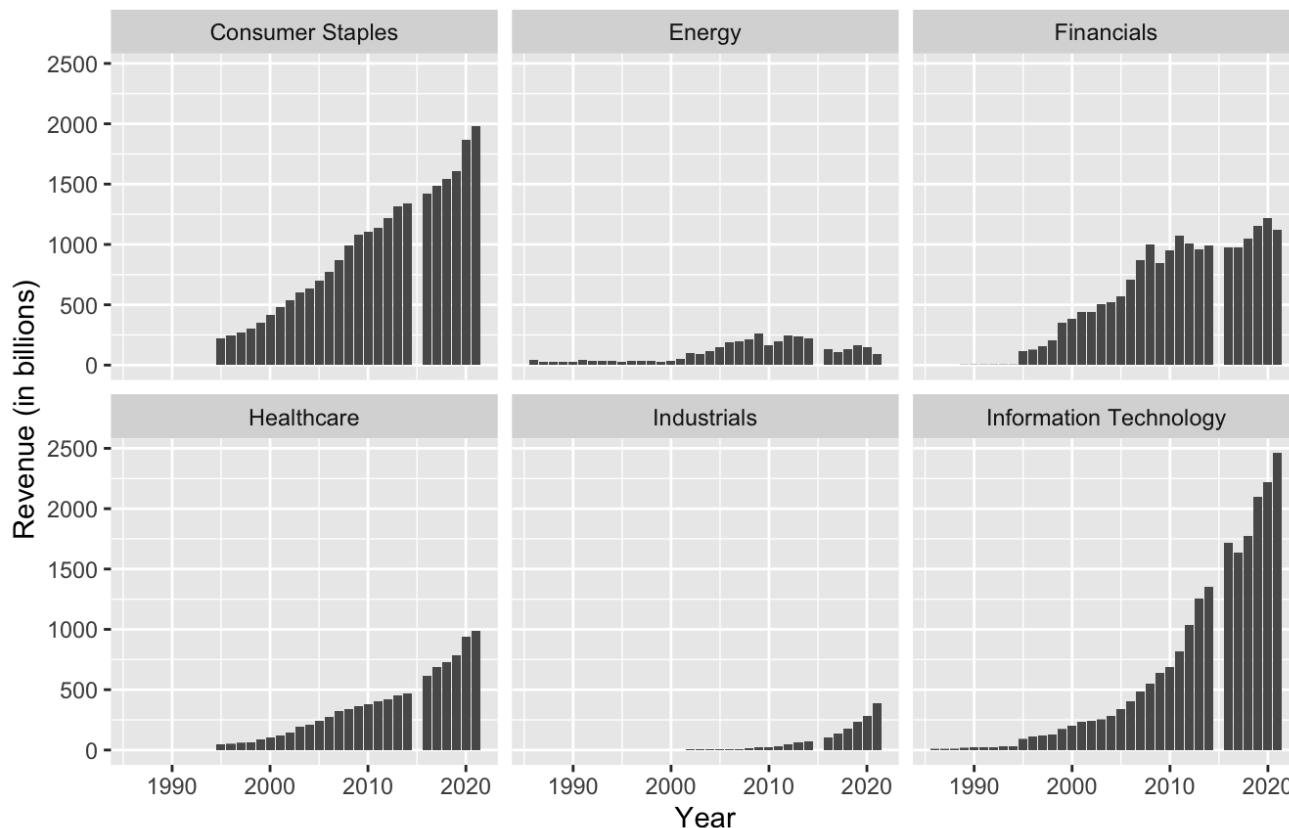
Displays distribution of Revenue, by company



```
# plotting all names, by sector
fortune_joined_v3 %>%
  filter(hyper_growth == "True") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col() +
  facet_wrap(~ Sector) +
  labs(title = "Revenue for Hyper-Growth Companies from 1985 to 2021",
       subtitle = "Displays distribution of Revenue, by Sector") +
  ylab("Revenue (in billions)")
```

Revenue for Hyper-Growth Companies from 1985 to 2021

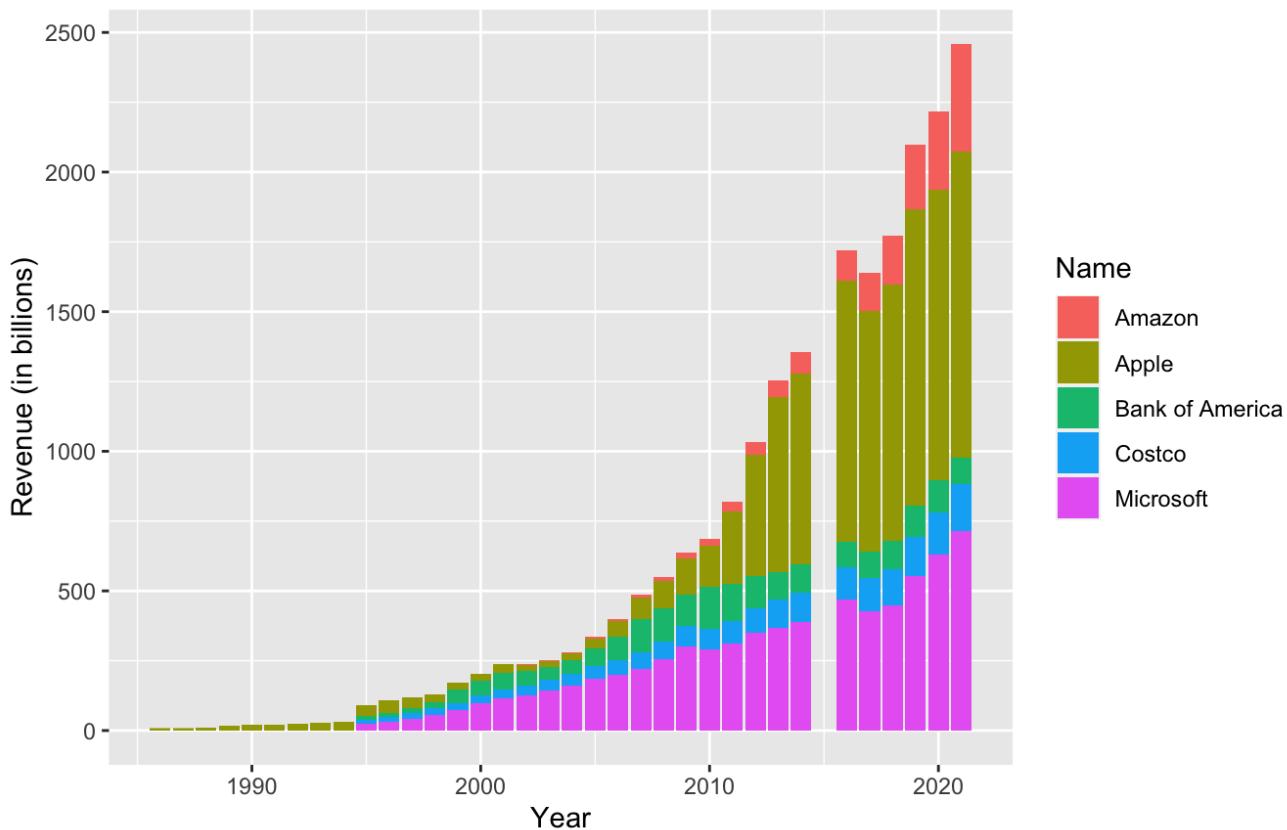
Displays distribution of Revenue, by Sector



```
# plotting IT companies
fortune_joined_v3 %>%
  filter(hyper_growth == "True", Sector == "Information Technology") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes(fill = Name)) +
  labs(title = "Revenue for Hyper-Growth Companies from 1985 to 2021",
       subtitle = "Information Technology Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Hyper-Growth Companies from 1985 to 2021

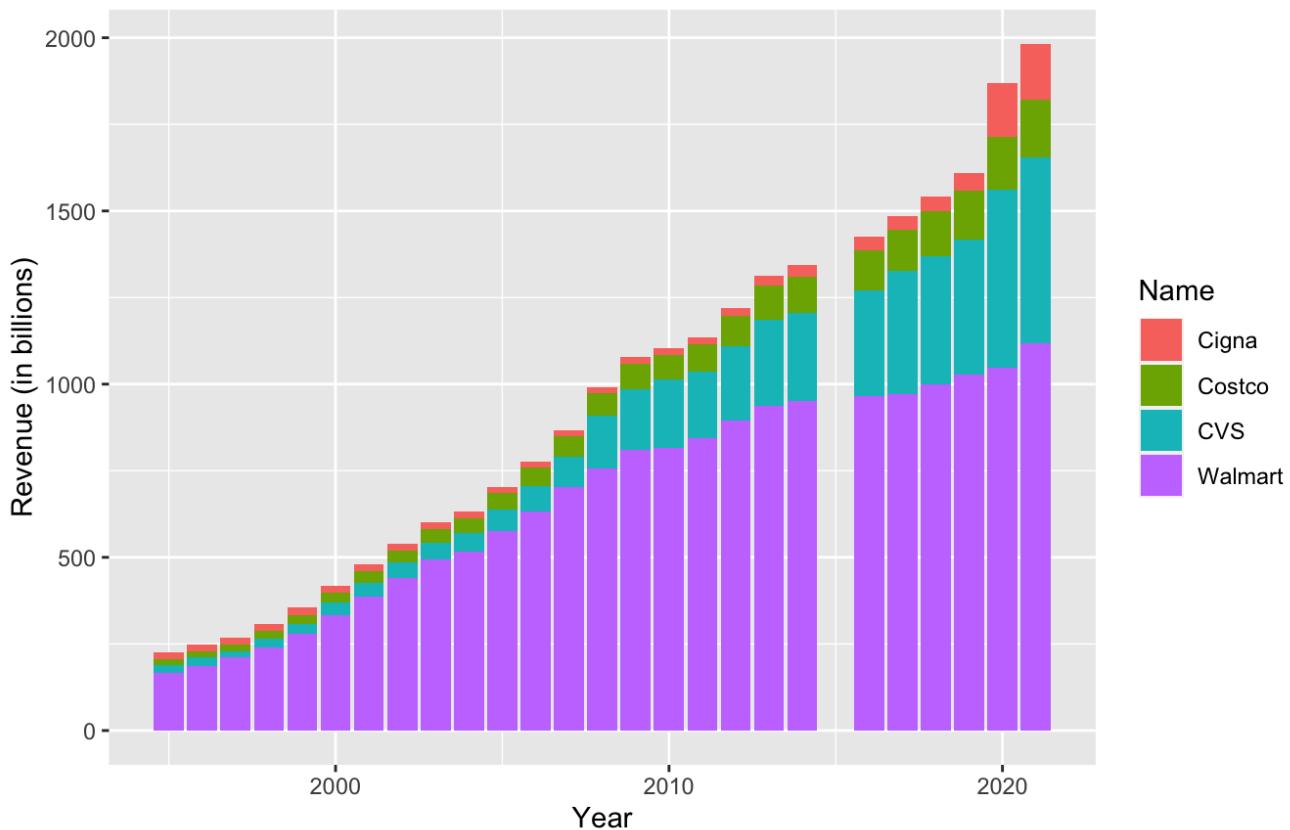
Information Technology Companies



```
# plotting Consumer Staples companies
fortune_joined_v3 %>%
  filter(hyper_growth == "True", Sector == "Consumer Staples") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes(fill = Name)) +
  labs(title = "Revenue for Hyper-Growth Companies from 1985 to 2021",
       subtitle = "Consumer Staples Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Hyper-Growth Companies from 1985 to 2021

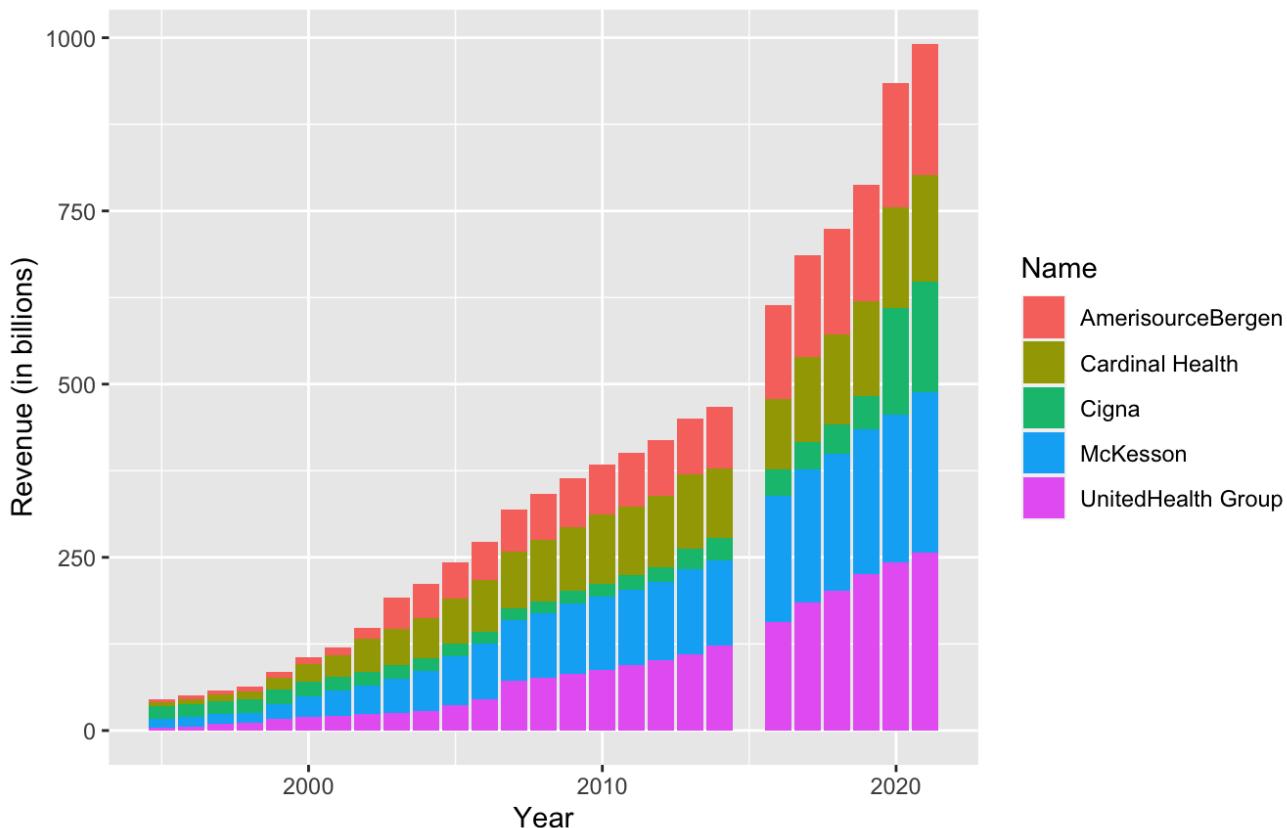
Consumer Staples Companies



```
# plotting Healthcare companies
fortune_joined_v3 %>%
  filter(hyper_growth == "True", Sector == "Healthcare") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes(fill = Name)) +
  labs(title = "Revenue for Hyper-Growth Companies from 1985 to 2021",
       subtitle = "Healthcare Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Hyper-Growth Companies from 1985 to 2021

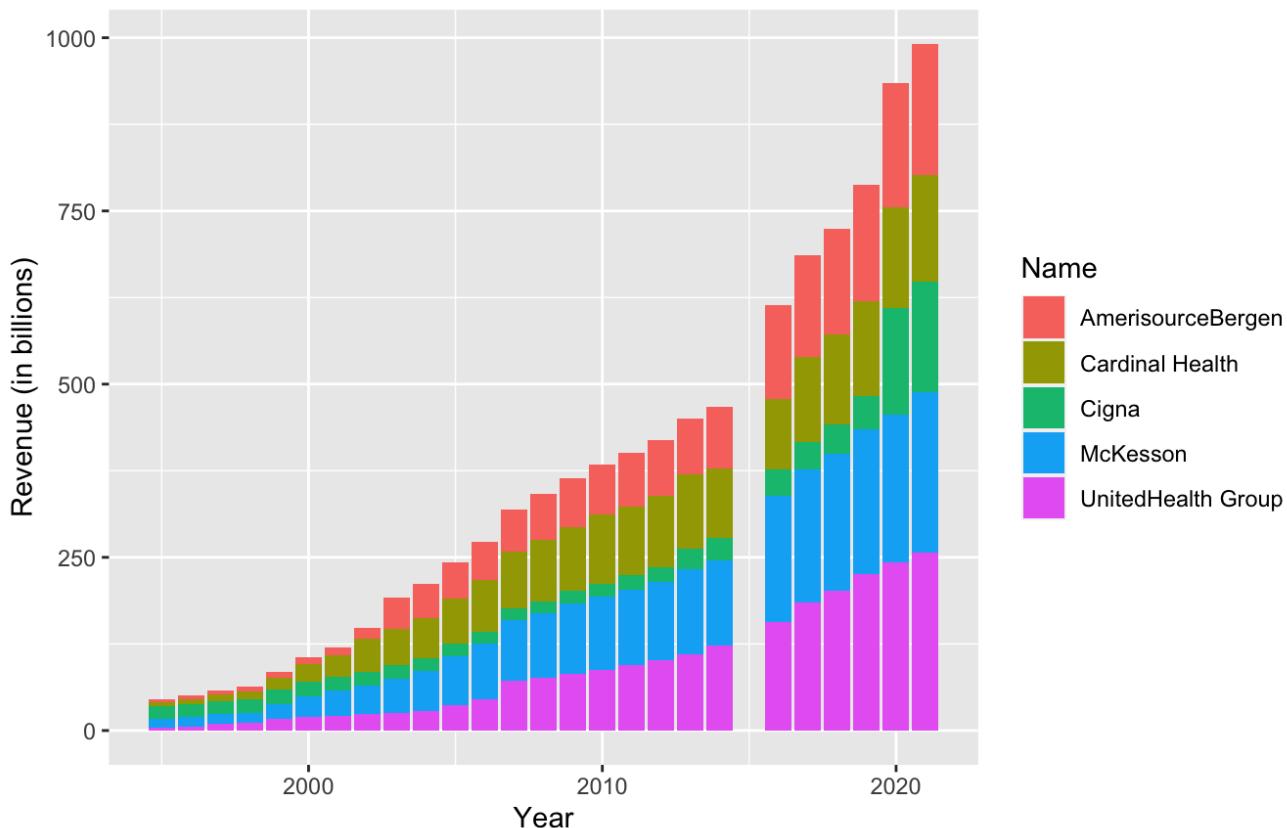
Healthcare Companies



```
# plotting Financials Companies
fortune_joined_v3 %>%
  filter(hyper_growth == "True", Sector == "Healthcare") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes(fill = Name)) +
  labs(title = "Revenue for Hyper-Growth Companies from 1985 to 2021",
       subtitle = "Financials Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Hyper-Growth Companies from 1985 to 2021

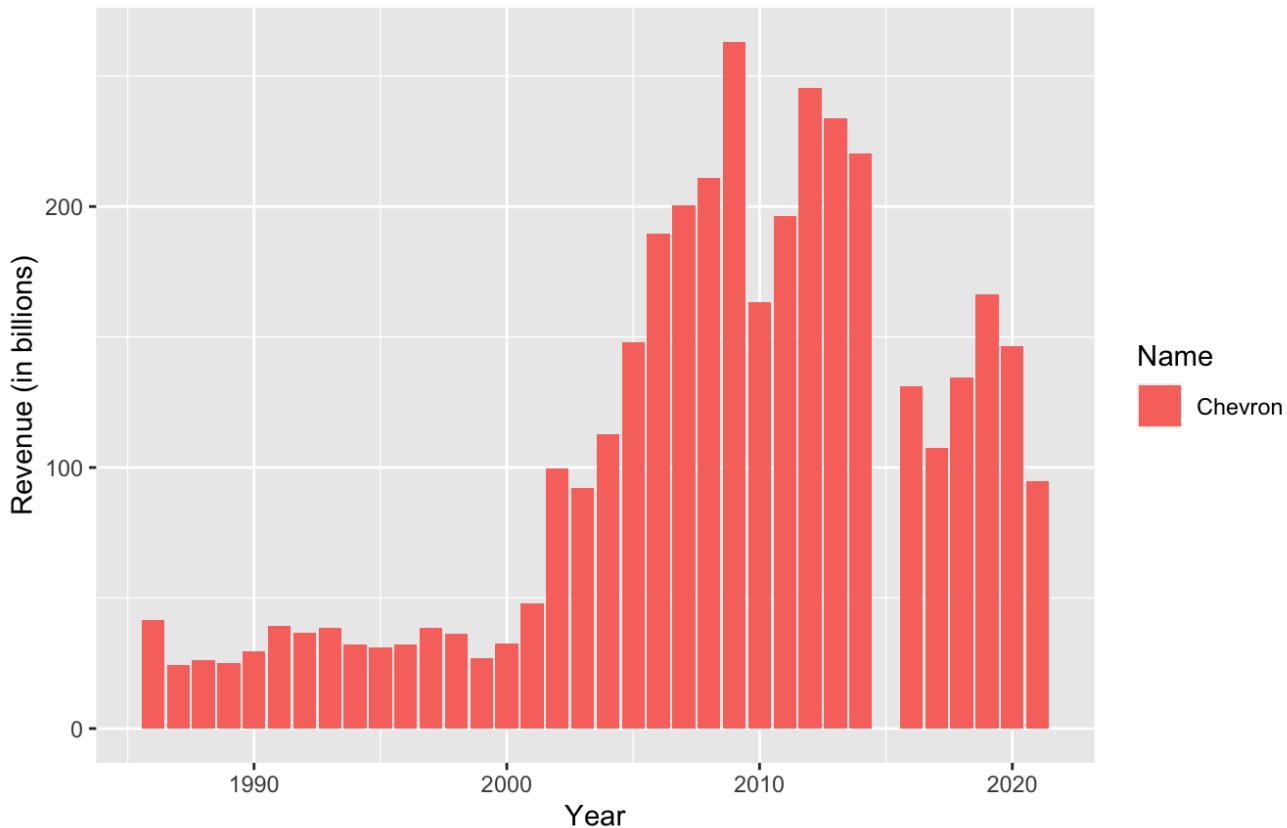
Financials Companies



```
# plotting Energy companies
fortune_joined_v3 %>%
  filter(hyper_growth == "True", Sector == "Energy") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes(fill = Name)) +
  labs(title = "Revenue for Hyper-Growth Companies from 1985 to 2021",
       subtitle = "Energy Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Hyper-Growth Companies from 1985 to 2021

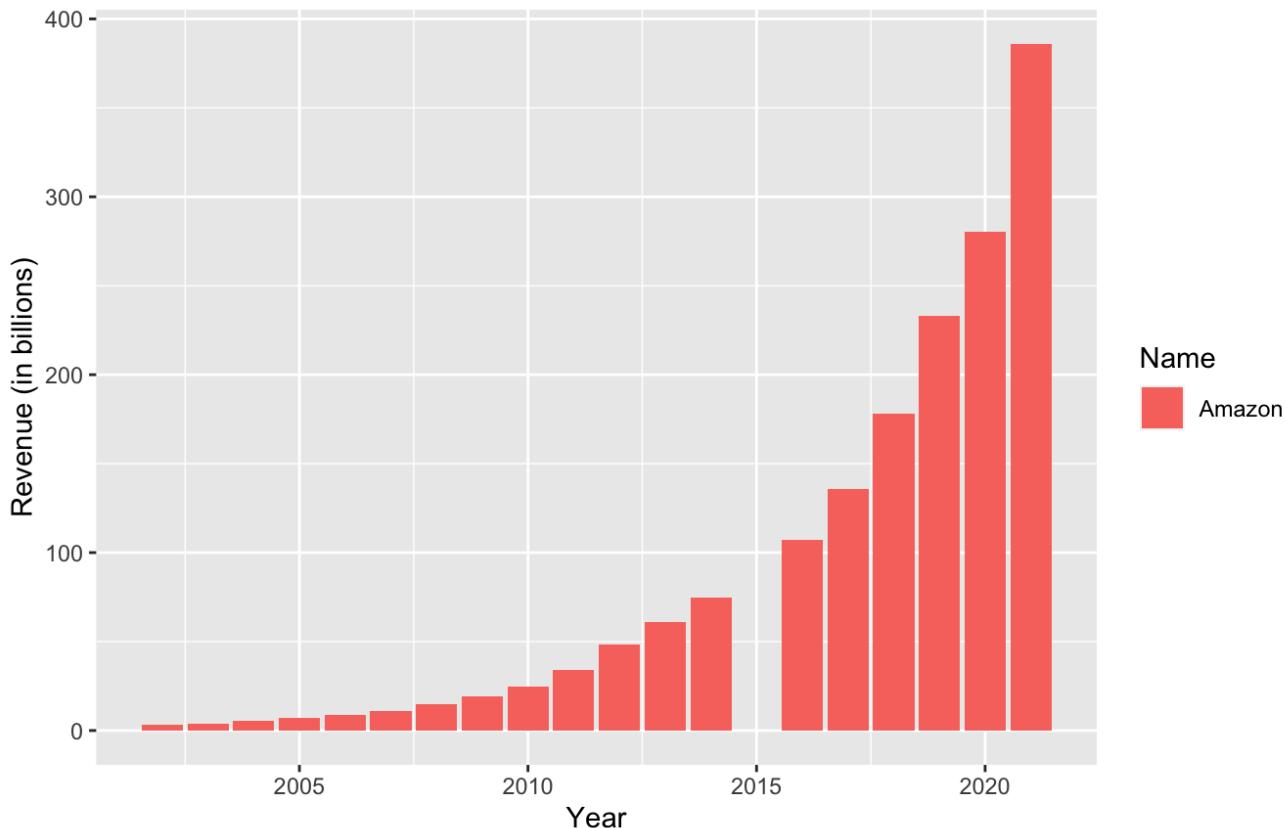
Energy Companies



```
# plotting Industrials companies
fortune_joined_v3 %>%
  filter(hyper_growth == "True", Sector == "Industrials") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes(fill = Name)) +
  labs(title = "Revenue for Hyper-Growth Companies from 1985 to 2021",
       subtitle = "Industrials Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Hyper-Growth Companies from 1985 to 2021

Industrials Companies



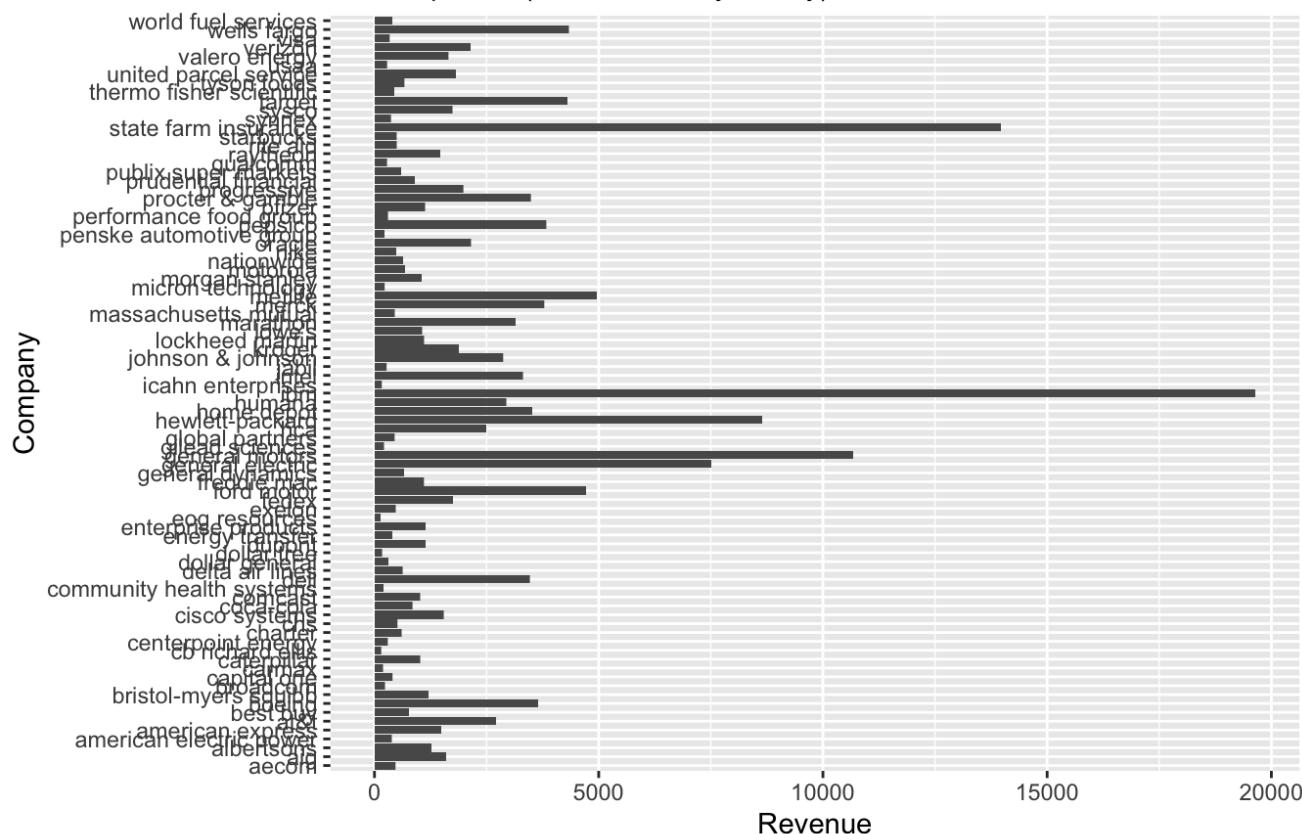
Moderate Growth Companies

As you can see, even at the moderate growth tier, it already becomes counterproductive to visualize all the companies.

```
# plotting all names
fortune_joined_v3 %>%
  filter(moderate_growth == "True") %>%
  ggplot(aes(Company, Revenue)) +
  geom_col() + coord_flip() +
  labs(title = "Revenue for Moderate-Growth Companies from 1985 to 2021",
       subtitle = "All Companies (understandably messy)")
```

Revenue for Moderate-Growth Companies from 1985 to 2021

All Companies (understandably messy)



It is more pragmatic to simply focus these companies on a sector-by-sector basis.

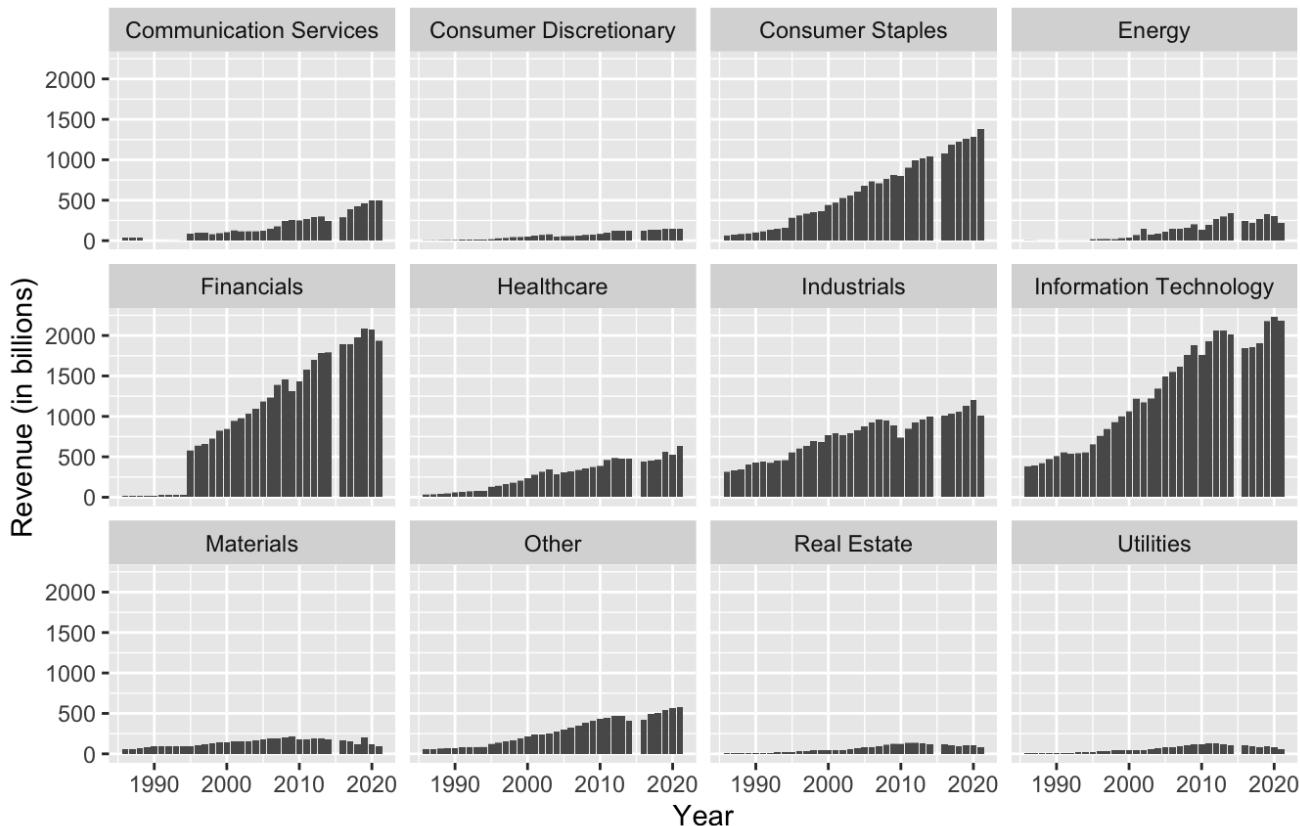
Observations:

- The consumer staples is less of a powerhouse in this selection.
- Moderate growth companies are still household names, and are certainly extremely successful, but they have grown at a much slower rate than the hyper-successful companies.

```
# plotting all companies, by sector
fortune_joined_v3 %>%
  filter(moderate_growth == "True") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col() +
  facet_wrap(~ Sector) +
  labs(title = "Revenue for Moderate-Growth Companies from 1985 to 2021",
       subtitle = "By Sector") +
  ylab("Revenue (in billions)")
```

Revenue for Moderate-Growth Companies from 1985 to 2021

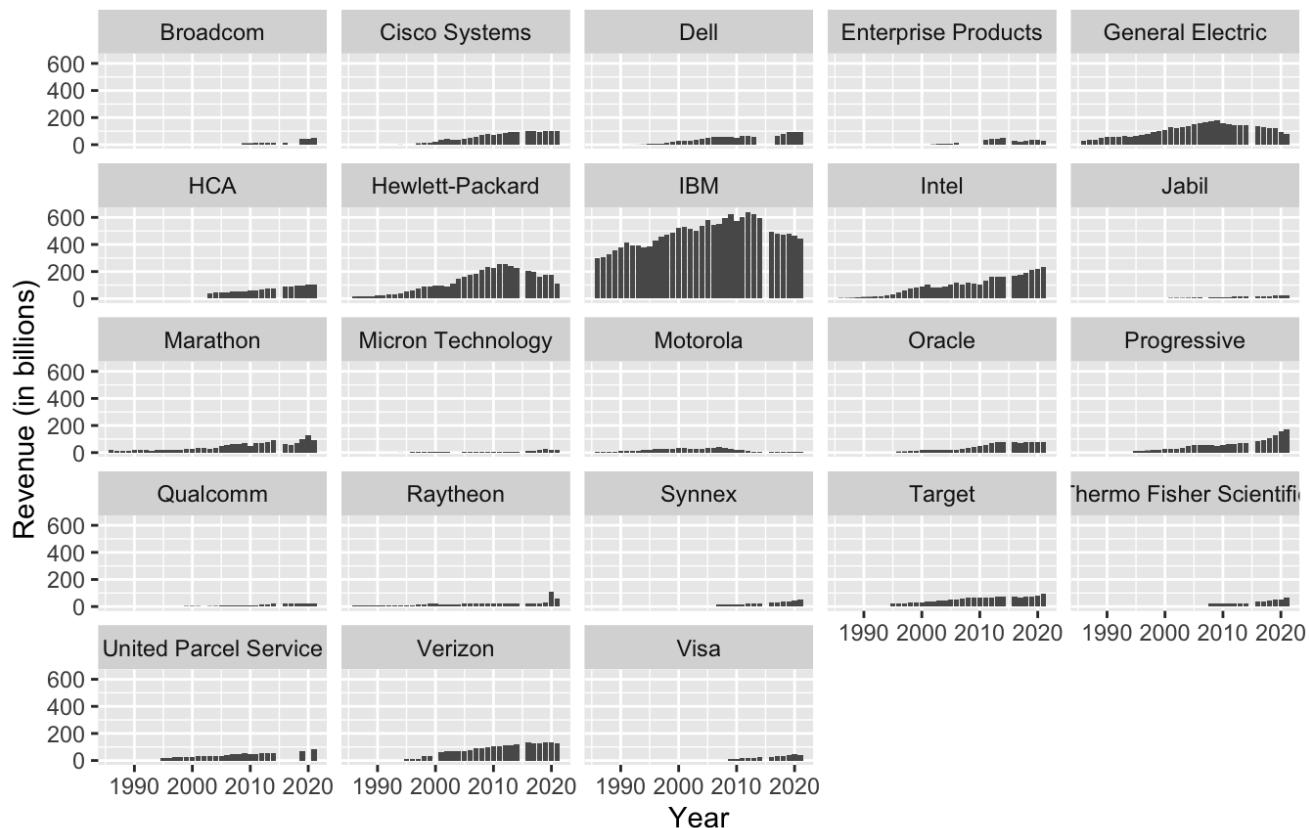
By Sector



```
# plotting IT companies
fortune_joined_v3 %>%
  filter(moderate_growth == "True", Sector == "Information Technology") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Moderate-Growth Companies from 1985 to 2021",
       subtitle = "Information Technology Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Moderate-Growth Companies from 1985 to 2021

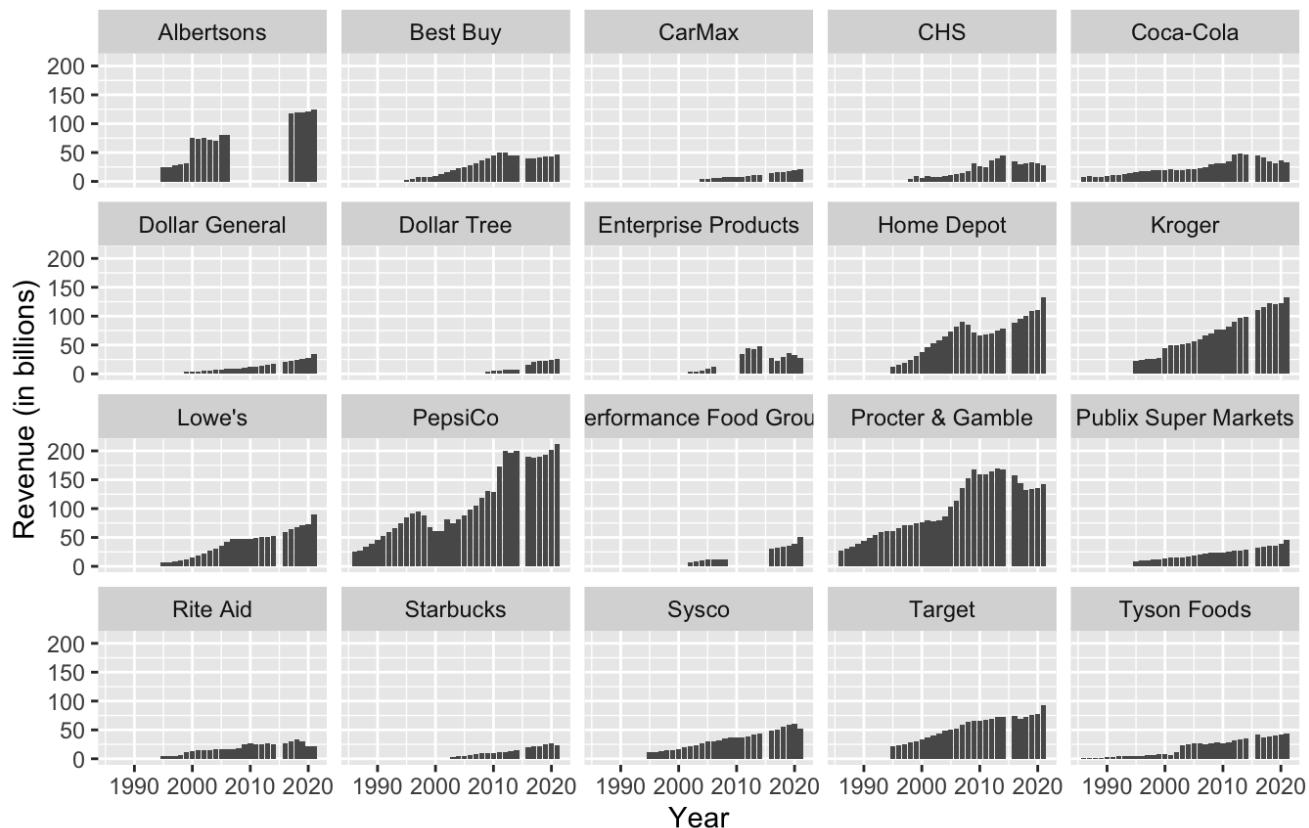
Information Technology Companies



```
# plotting Consumer Staples companies
fortune_joined_v3 %>%
  filter(moderate_growth == "True", Sector == "Consumer Staples") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Moderate-Growth Companies from 1985 to 2021",
       subtitle = "Consumer Staples Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Moderate-Growth Companies from 1985 to 2021

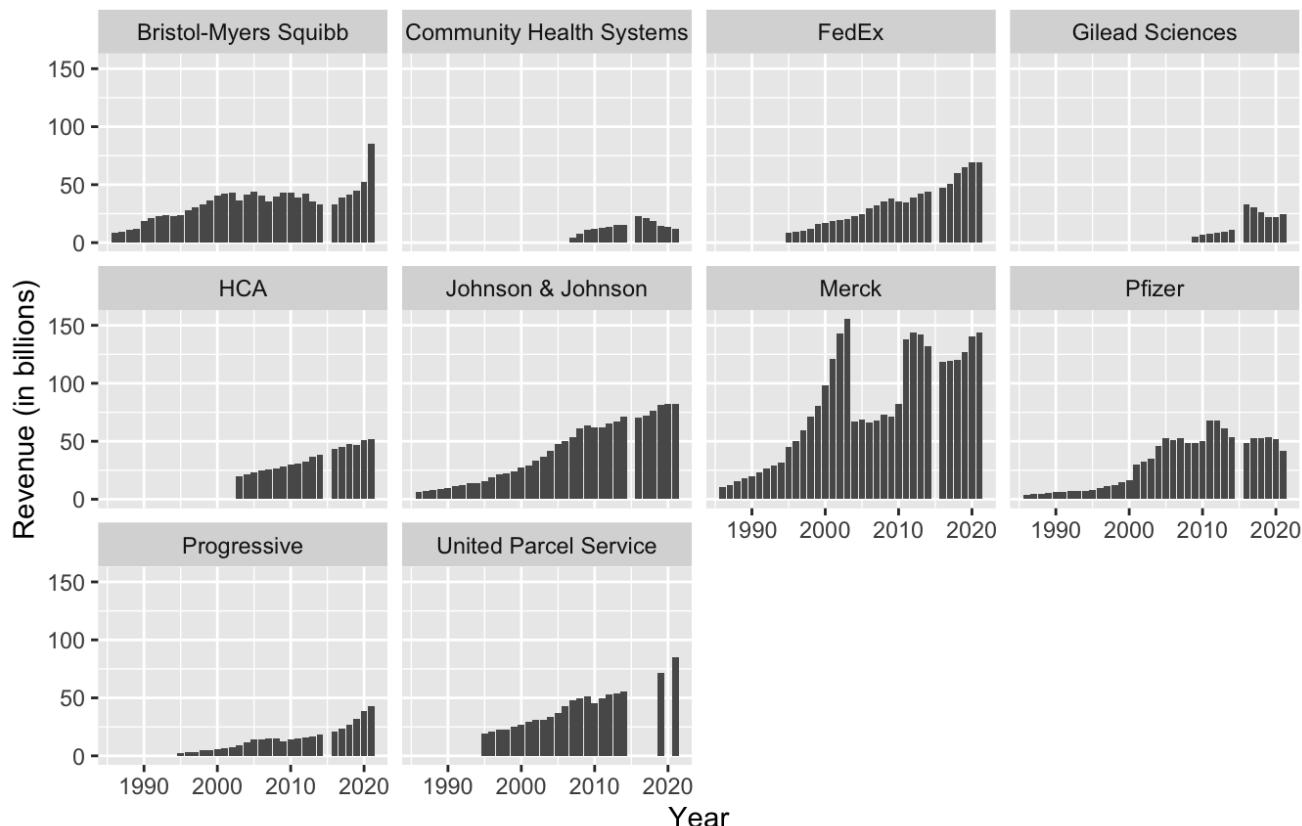
Consumer Staples Companies



```
# plotting Healthcare companies
fortune_joined_v3 %>%
  filter(moderate_growth == "True", Sector == "Healthcare") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Moderate-Growth Companies from 1985 to 2021",
       subtitle = "Healthcare Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Moderate-Growth Companies from 1985 to 2021

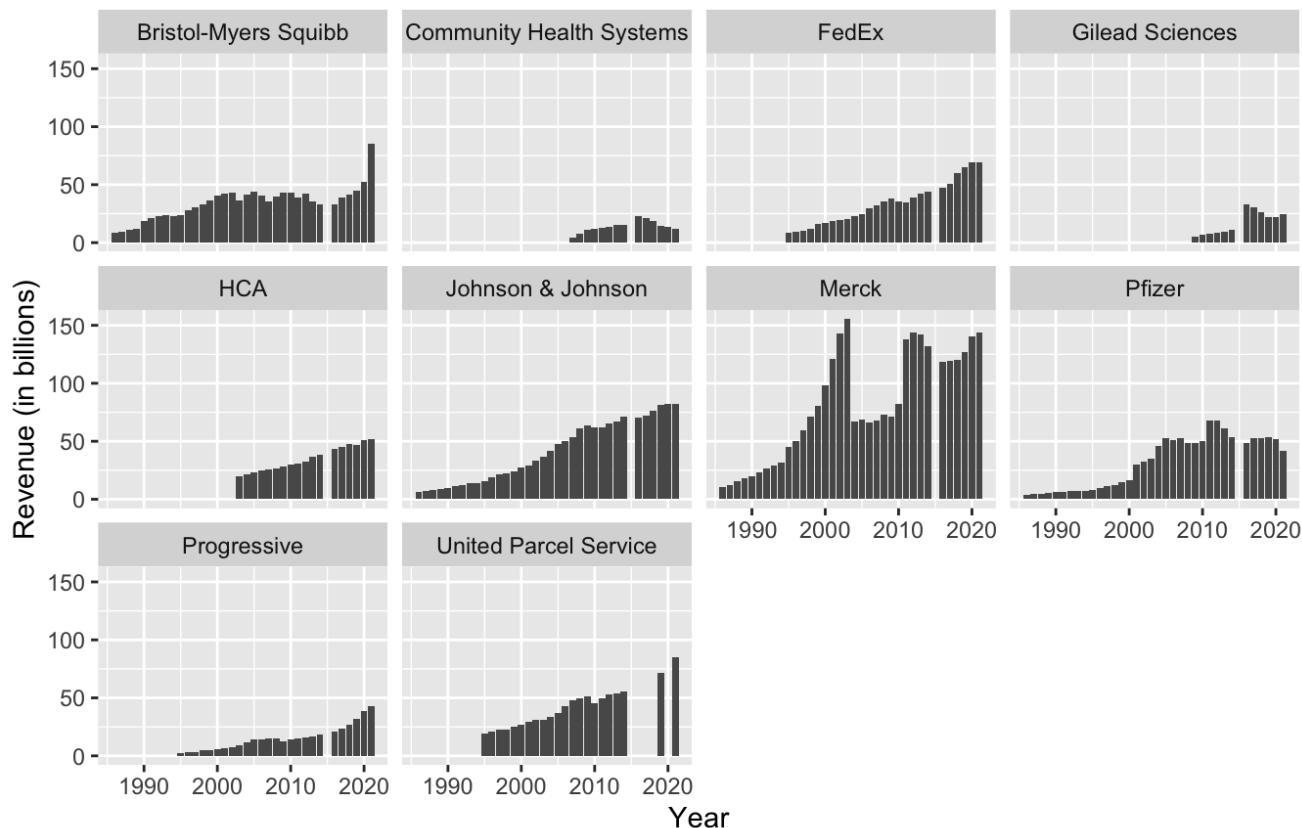
Healthcare Companies



```
# plotting Financials Companies
fortune_joined_v3 %>%
  filter(moderate_growth == "True", Sector == "Healthcare") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Moderate-Growth Companies from 1985 to 2021",
       subtitle = "Financials Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Moderate-Growth Companies from 1985 to 2021

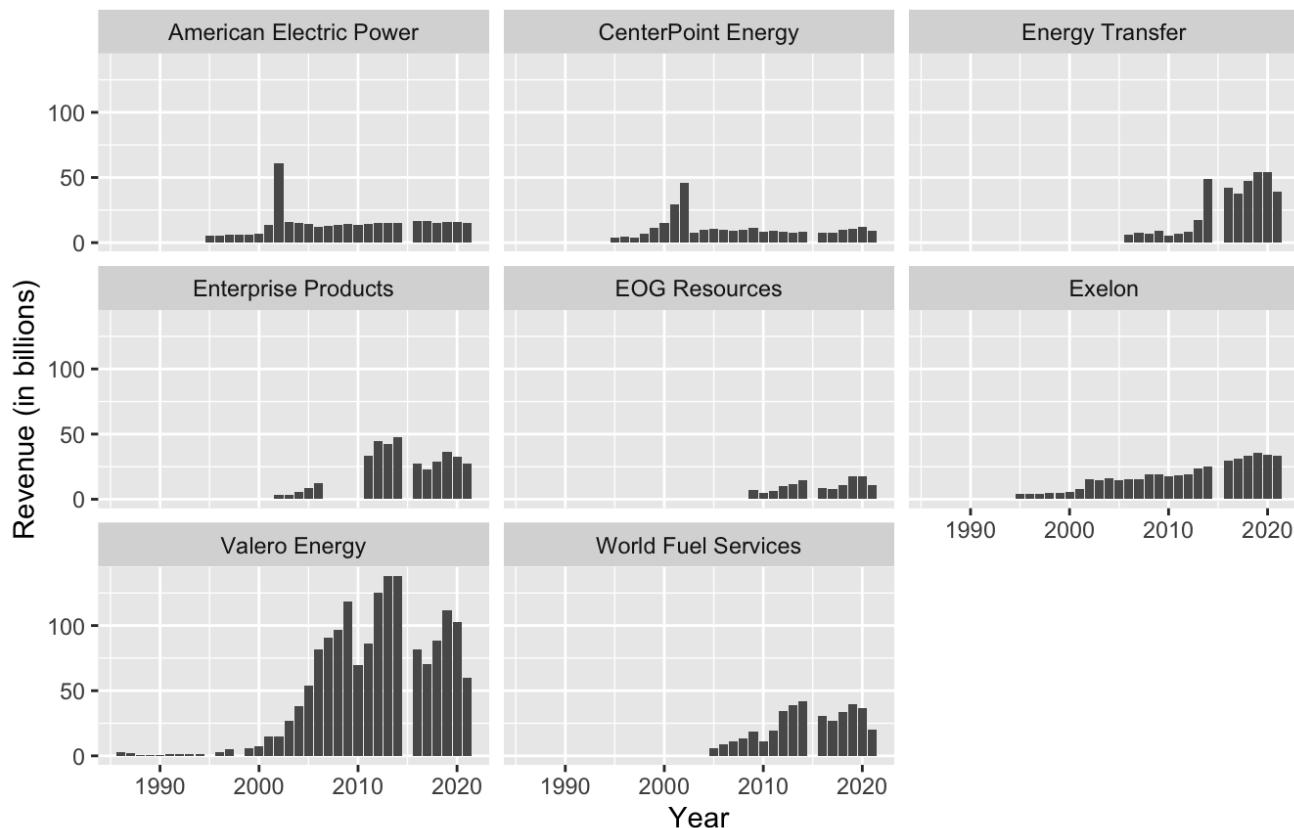
Financials Companies



```
# plotting Energy companies
fortune_joined_v3 %>%
  filter(moderate_growth == "True", Sector == "Energy") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Moderate-Growth Companies from 1985 to 2021",
       subtitle = "Energy Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Moderate-Growth Companies from 1985 to 2021

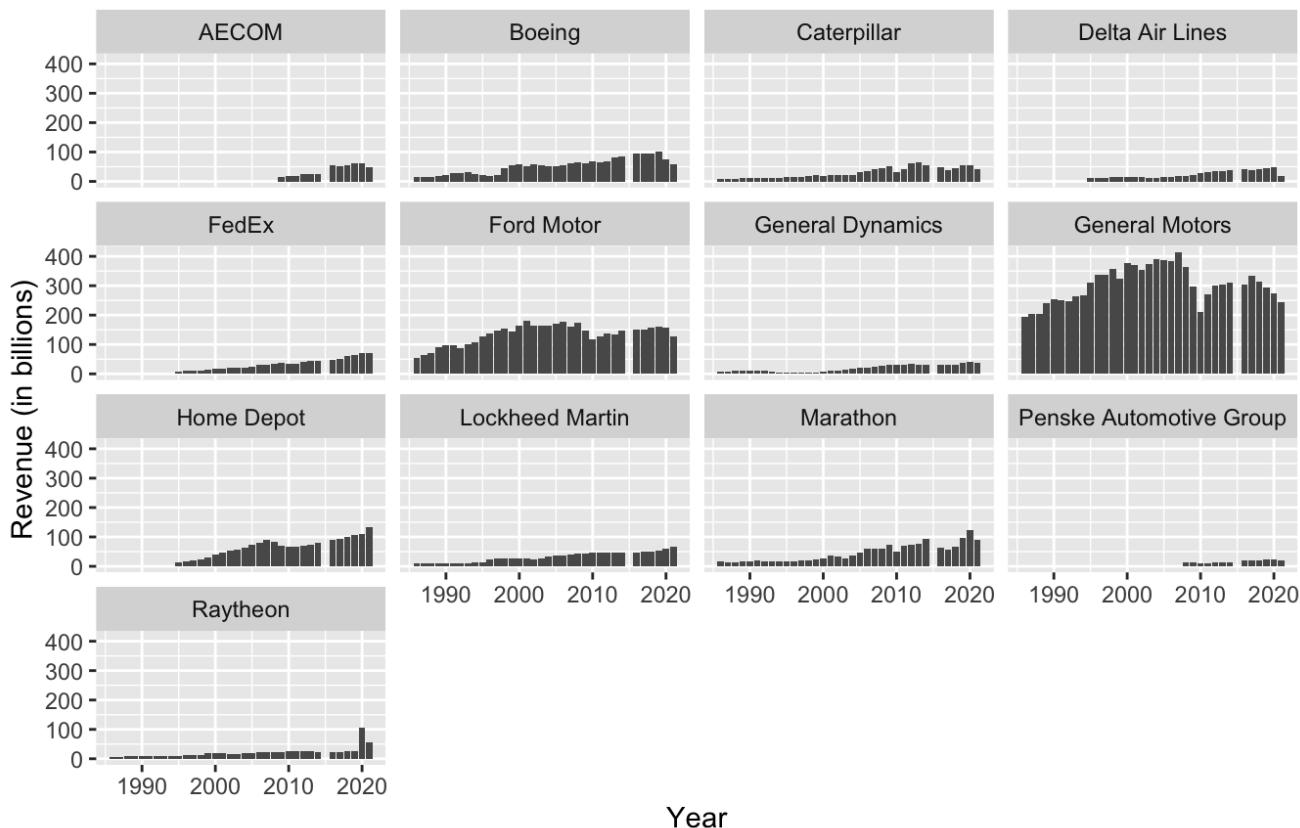
Energy Companies



```
# plotting Industrials companies
fortune_joined_v3 %>%
  filter(moderate_growth == "True", Sector == "Industrials") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Moderate-Growth Companies from 1985 to 2021",
       subtitle = "Industrials Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Moderate-Growth Companies from 1985 to 2021

Industrials Companies



Low Growth Companies

It makes more sense to simply focus these companies on a sector-by-sector basis.

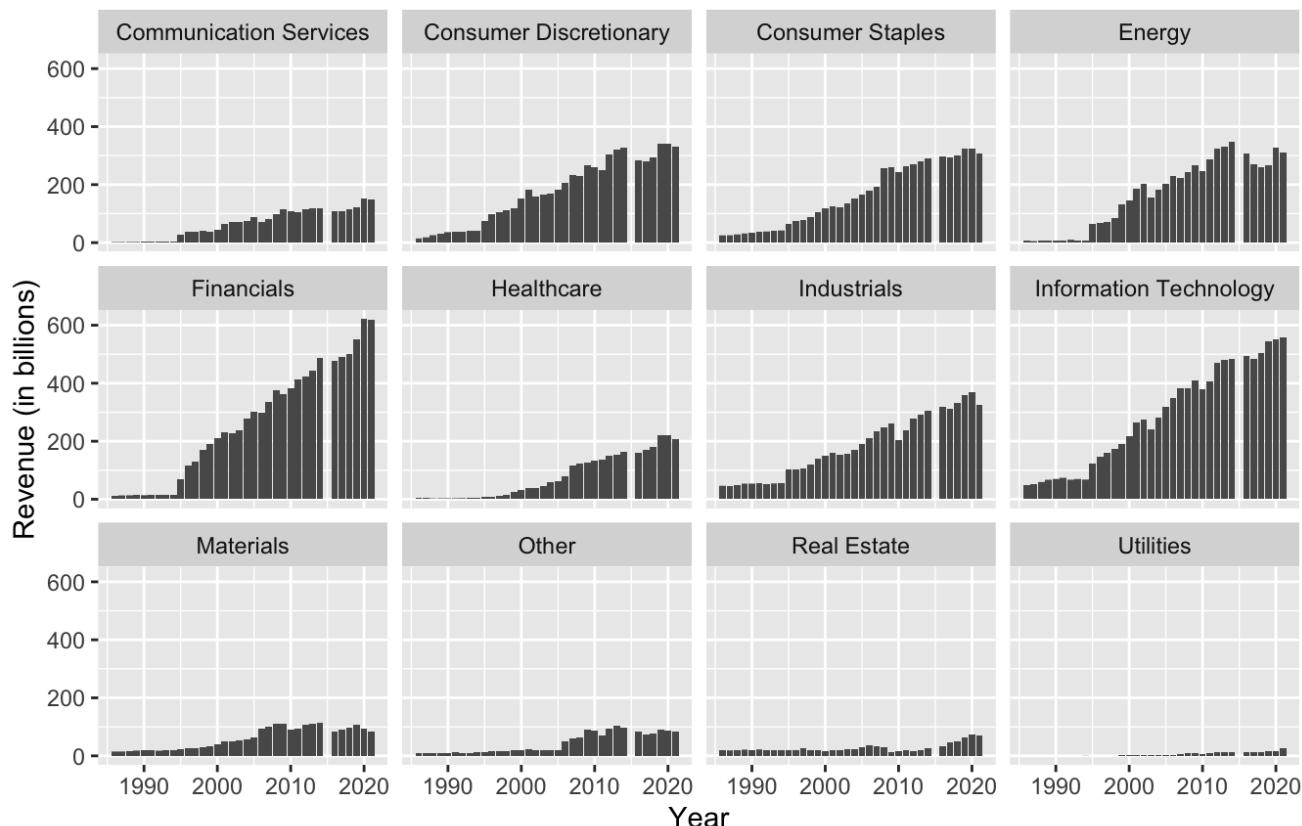
Observations:

- Low growth exhibit continuous growth, but at a much lower rate than that of the hyper-successful companies.
- It starts to become pretty noticeable that these companies have not maintained standing on the Fortune 500 every year.

```
# plotting all companies, by sector
fortune_joined_v3 %>%
  filter(low_growth == "True") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col() +
  facet_wrap(~ Sector) +
  labs(title = "Revenue for Low-Growth Companies from 1985 to 2021",
       subtitle = "By Sector") +
  ylab("Revenue (in billions)")
```

Revenue for Low-Growth Companies from 1985 to 2021

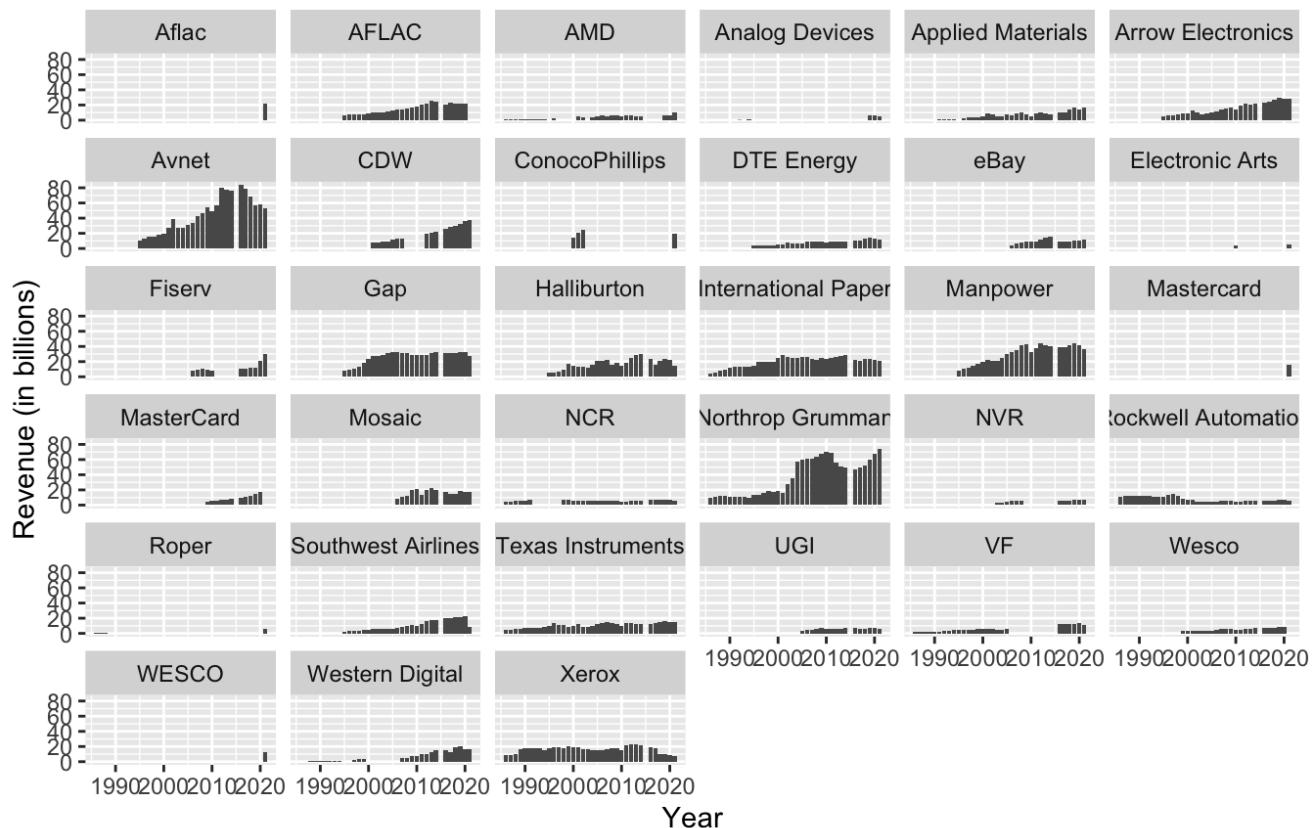
By Sector



```
# plotting IT companies
fortune_joined_v3 %>%
  filter(low_growth == "True", Sector == "Information Technology") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Low-Growth Companies from 1985 to 2021",
       subtitle = "Information Technology Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Low-Growth Companies from 1985 to 2021

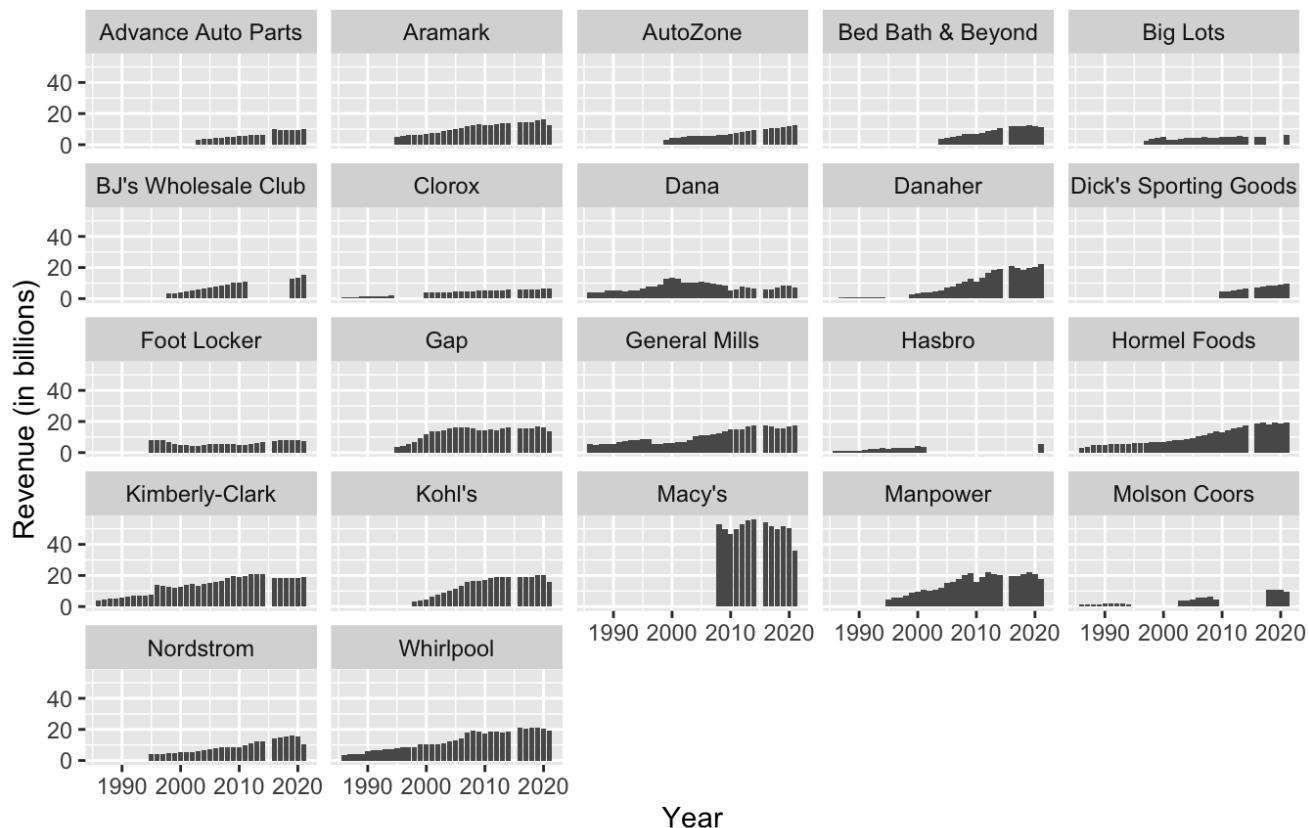
Information Technology Companies



```
# plotting Consumer Staples companies
fortune_joined_v3 %>%
  filter(low_growth == "True", Sector == "Consumer Staples") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Low-Growth Companies from 1985 to 2021",
       subtitle = "Consumer Staples Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Low-Growth Companies from 1985 to 2021

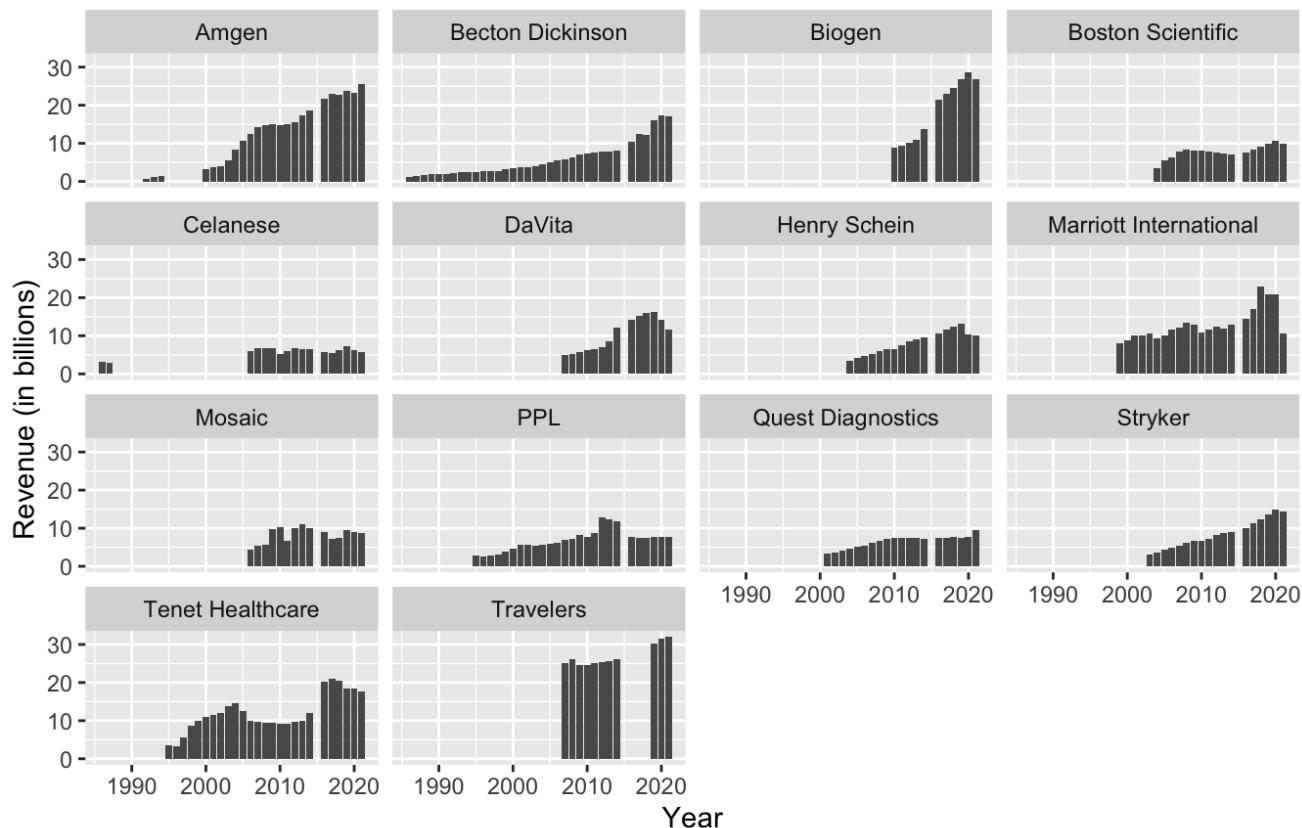
Consumer Staples Companies



```
# plotting Healthcare companies
fortune_joined_v3 %>%
  filter(low_growth == "True", Sector == "Healthcare") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Low-Growth Companies from 1985 to 2021",
       subtitle = "Healthcare Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Low-Growth Companies from 1985 to 2021

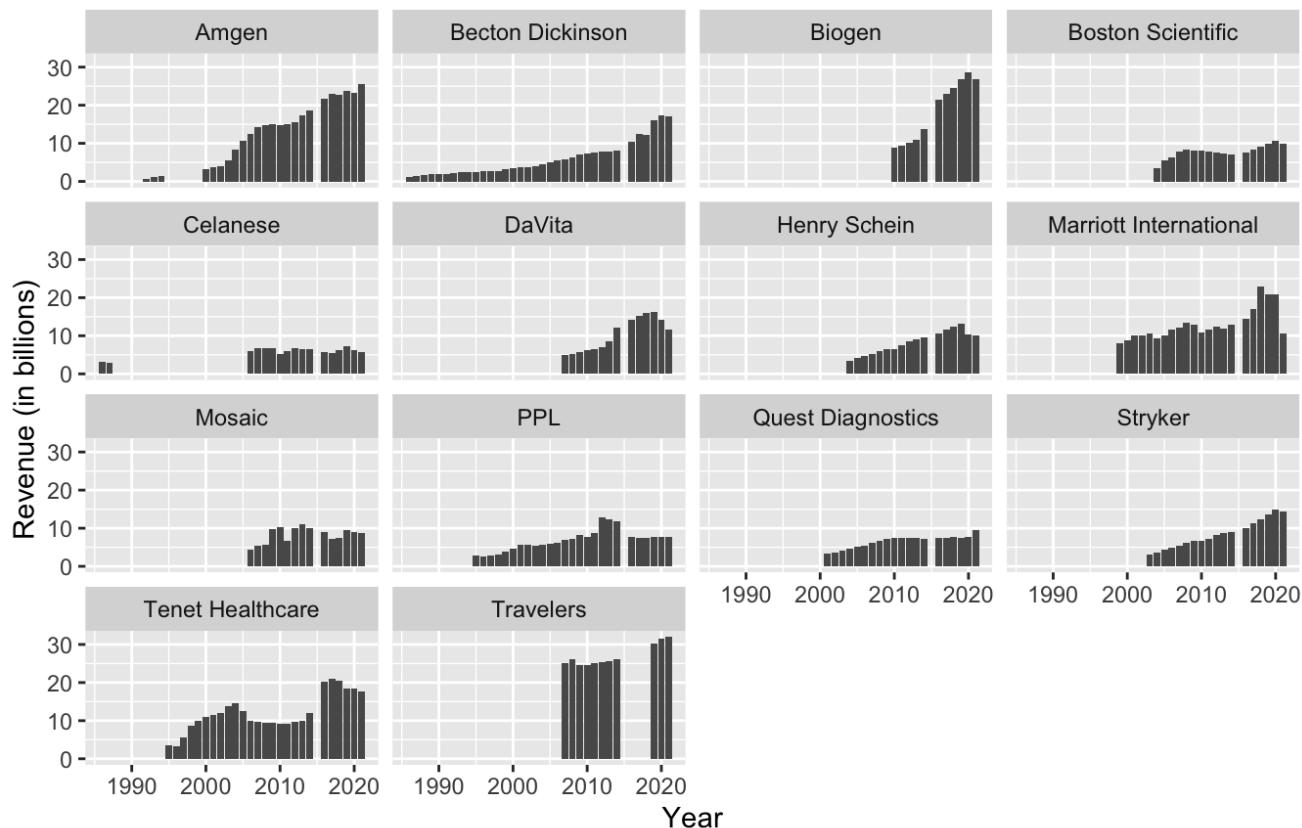
Healthcare Companies



```
# plotting Financials Companies
fortune_joined_v3 %>%
  filter(low_growth == "True", Sector == "Healthcare") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Low-Growth Companies from 1985 to 2021",
       subtitle = "Financials Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Low-Growth Companies from 1985 to 2021

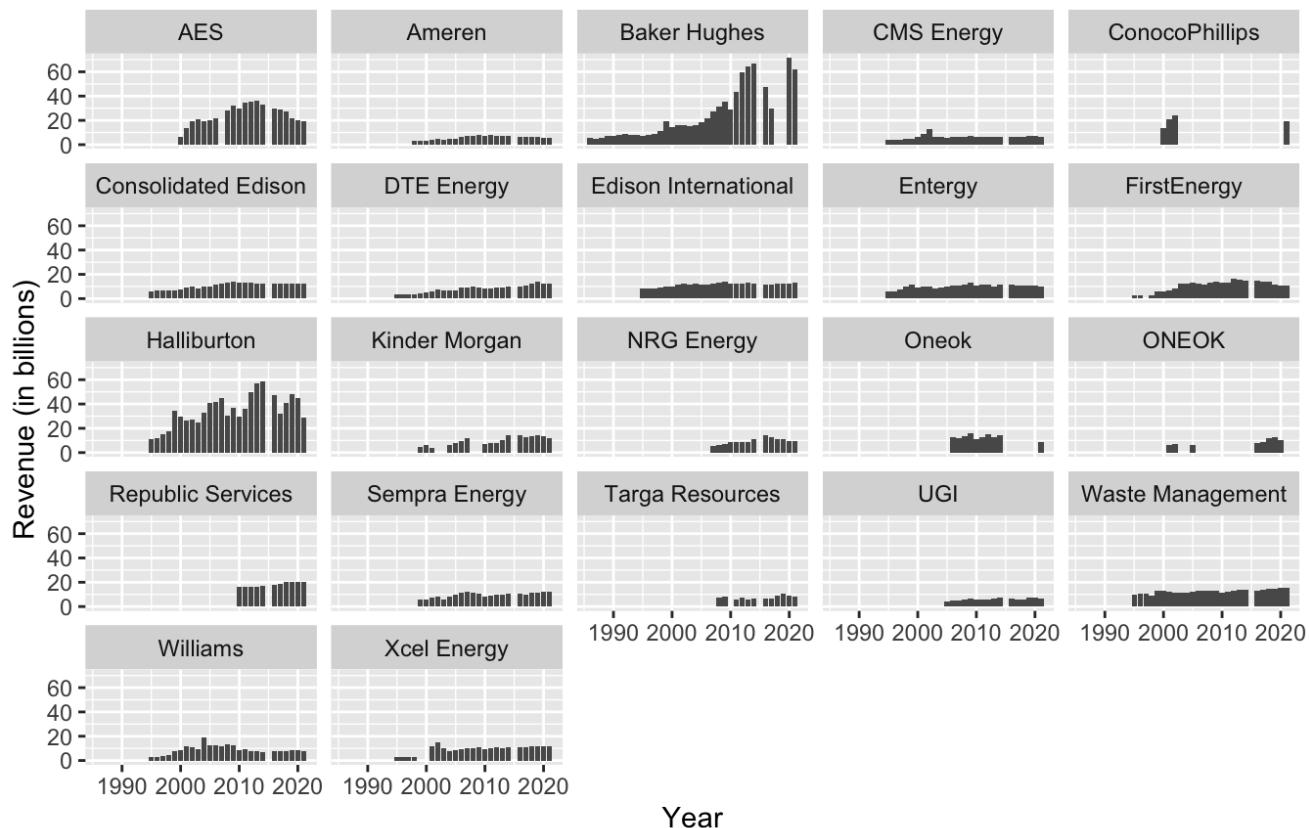
Financials Companies



```
# plotting Energy companies
fortune_joined_v3 %>%
  filter(low_growth == "True", Sector == "Energy") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Low-Growth Companies from 1985 to 2021",
       subtitle = "Energy Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Low-Growth Companies from 1985 to 2021

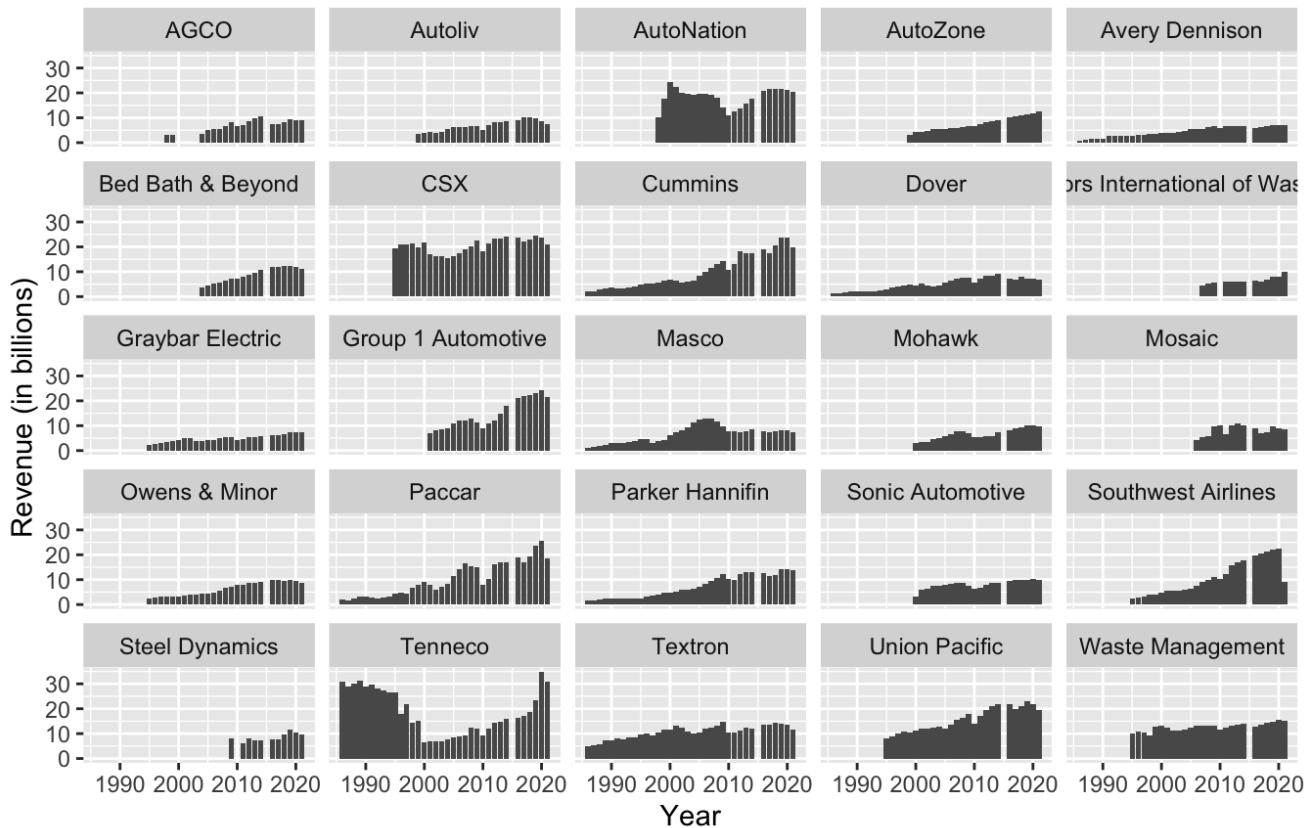
Energy Companies



```
# plotting Industrials companies
fortune_joined_v3 %>%
  filter(low_growth == "True", Sector == "Industrials") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Low-Growth Companies from 1985 to 2021",
       subtitle = "Industrials Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Low-Growth Companies from 1985 to 2021

Industrials Companies



Unsuccessful Companies

Again, for these companies it makes much more sense to analyze them on a sector by sector basis.

*Keep in mind, these are still very successful companies, this is simply relative to the extremely successful ones.

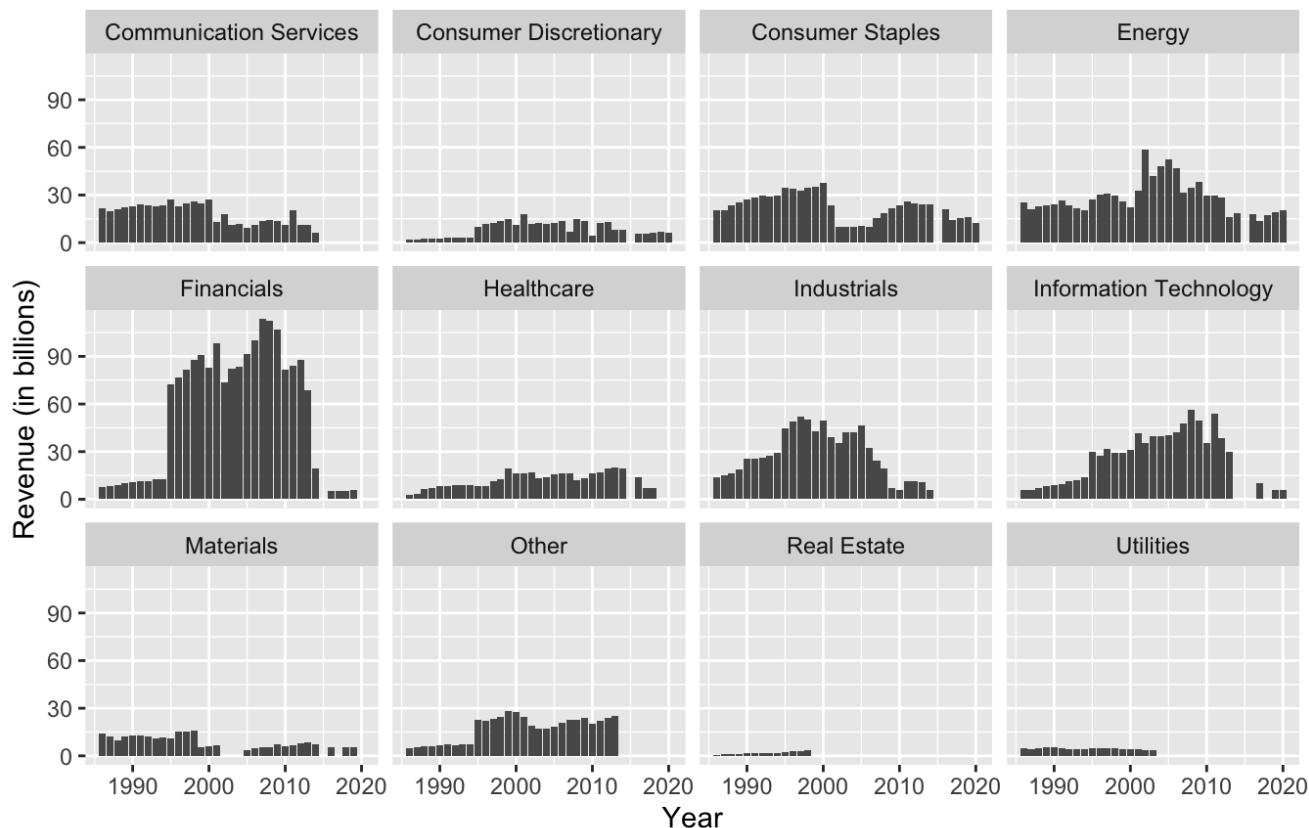
Observations:

- The plots for revenue for unsuccessful companies behave as expected, with little growth, and long term lack thereof.
- Unsurprisingly, the unsuccessful companies have much less commonly-heard names, which indicates that they may not be as good at using technological advances such as the internet for efficient distribution of their products.

```
# plotting all companies, by sector
fortune_joined_v3 %>%
  filter(unsuccessful == "True") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col() +
  facet_wrap(~ Sector) +
  labs(title = "Revenue for Unsuccessful* Companies from 1985 to 2021",
       subtitle = "By Sector") +
  ylab("Revenue (in billions)")
```

Revenue for Unsuccessful* Companies from 1985 to 2021

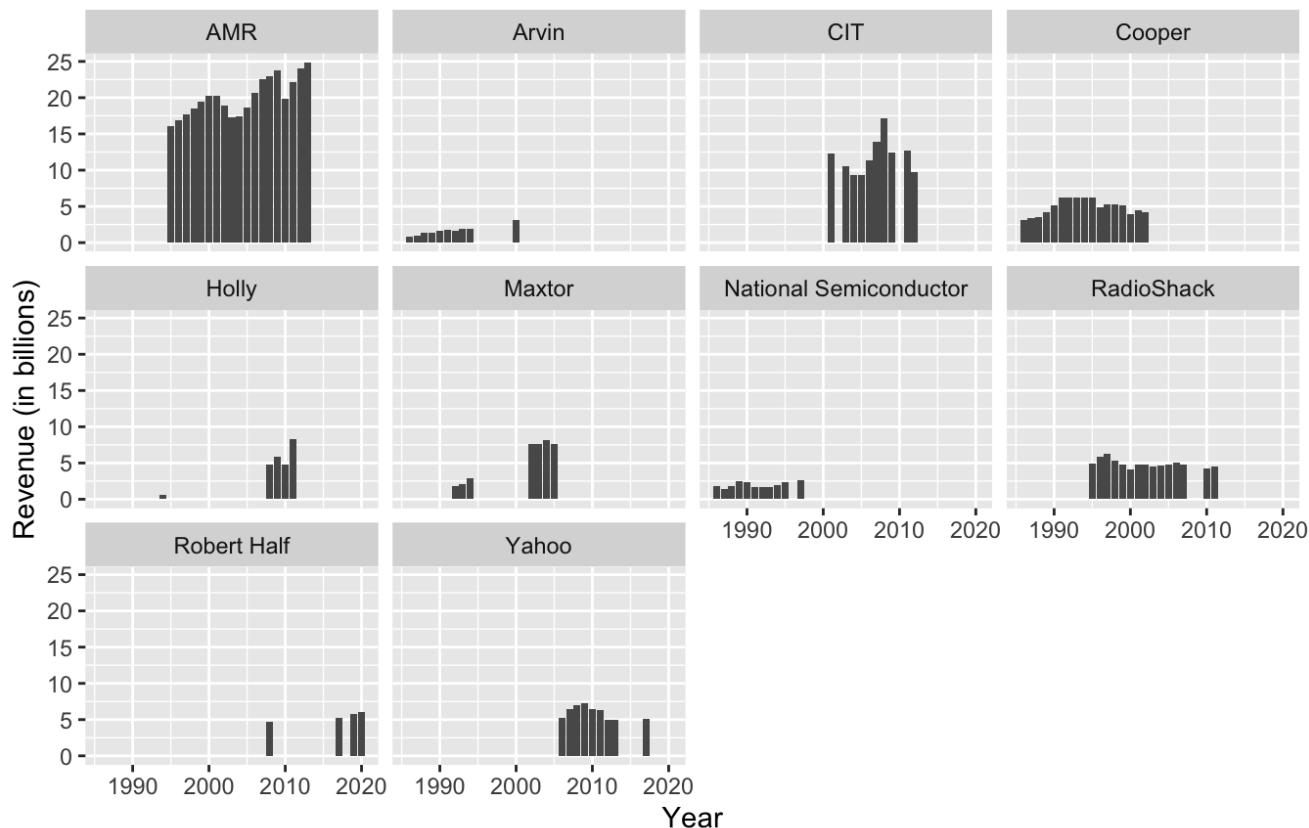
By Sector



```
# plotting IT companies
fortune_joined_v3 %>%
  filter(unsuccessful == "True", Sector == "Information Technology") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Unsuccessful* Companies from 1985 to 2021",
       subtitle = "Information Technology Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Unsuccessful* Companies from 1985 to 2021

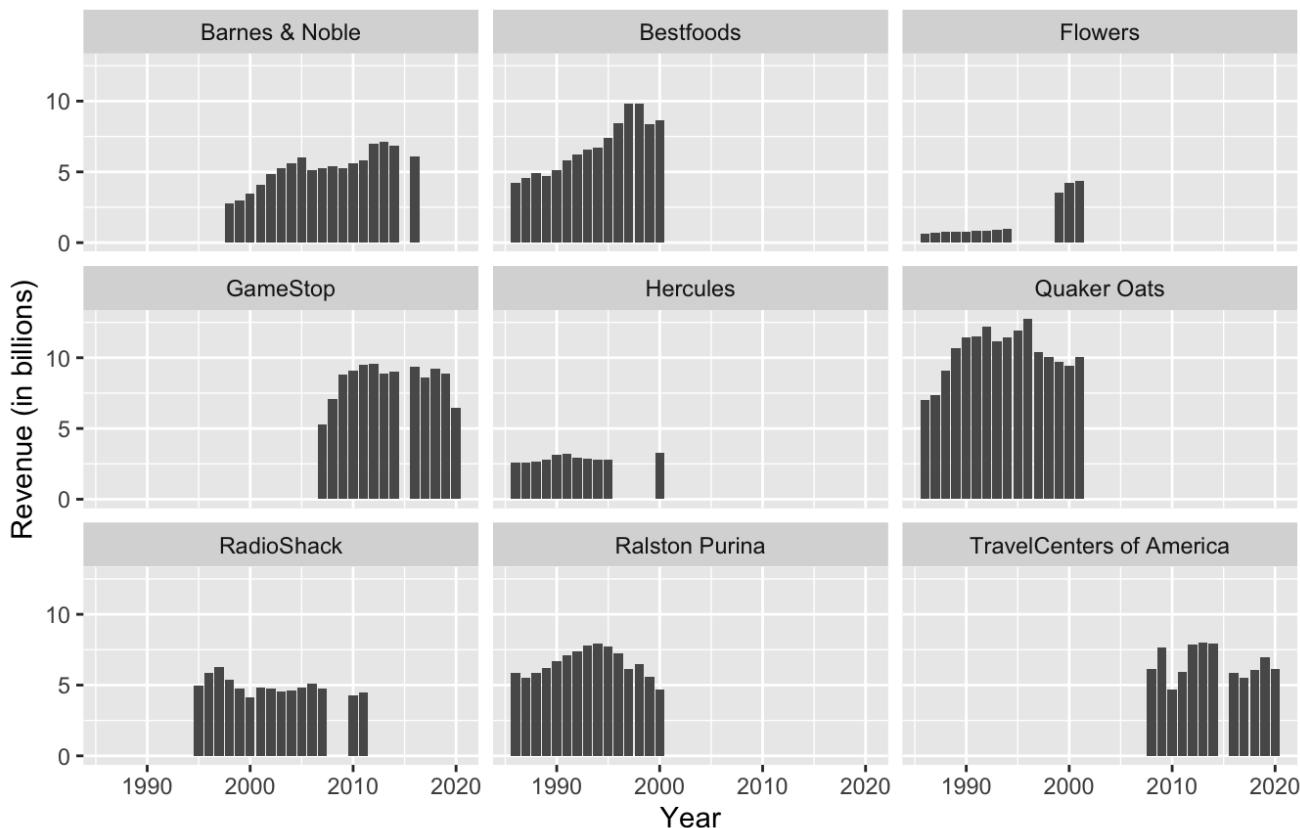
Information Technology Companies



```
# plotting Consumer Staples companies
fortune_joined_v3 %>%
  filter(unsuccessful == "True", Sector == "Consumer Staples") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Unsuccessful* Companies from 1985 to 2021",
       subtitle = "Consumer Staples Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Unsuccessful* Companies from 1985 to 2021

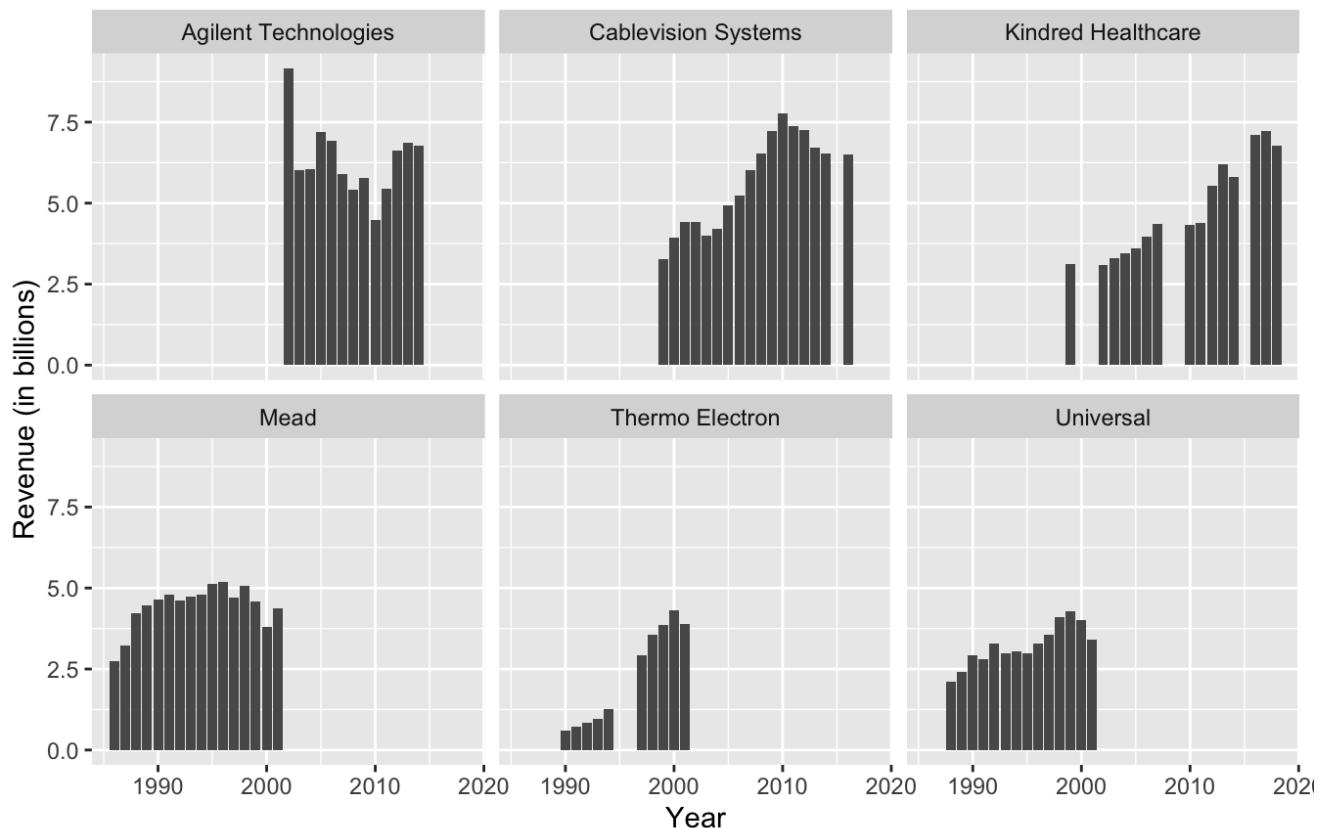
Consumer Staples Companies



```
# plotting Healthcare companies
fortune_joined_v3 %>%
  filter(unsuccessful == "True", Sector == "Healthcare") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Unsuccessful* Companies from 1985 to 2021",
       subtitle = "Healthcare Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Unsuccessful* Companies from 1985 to 2021

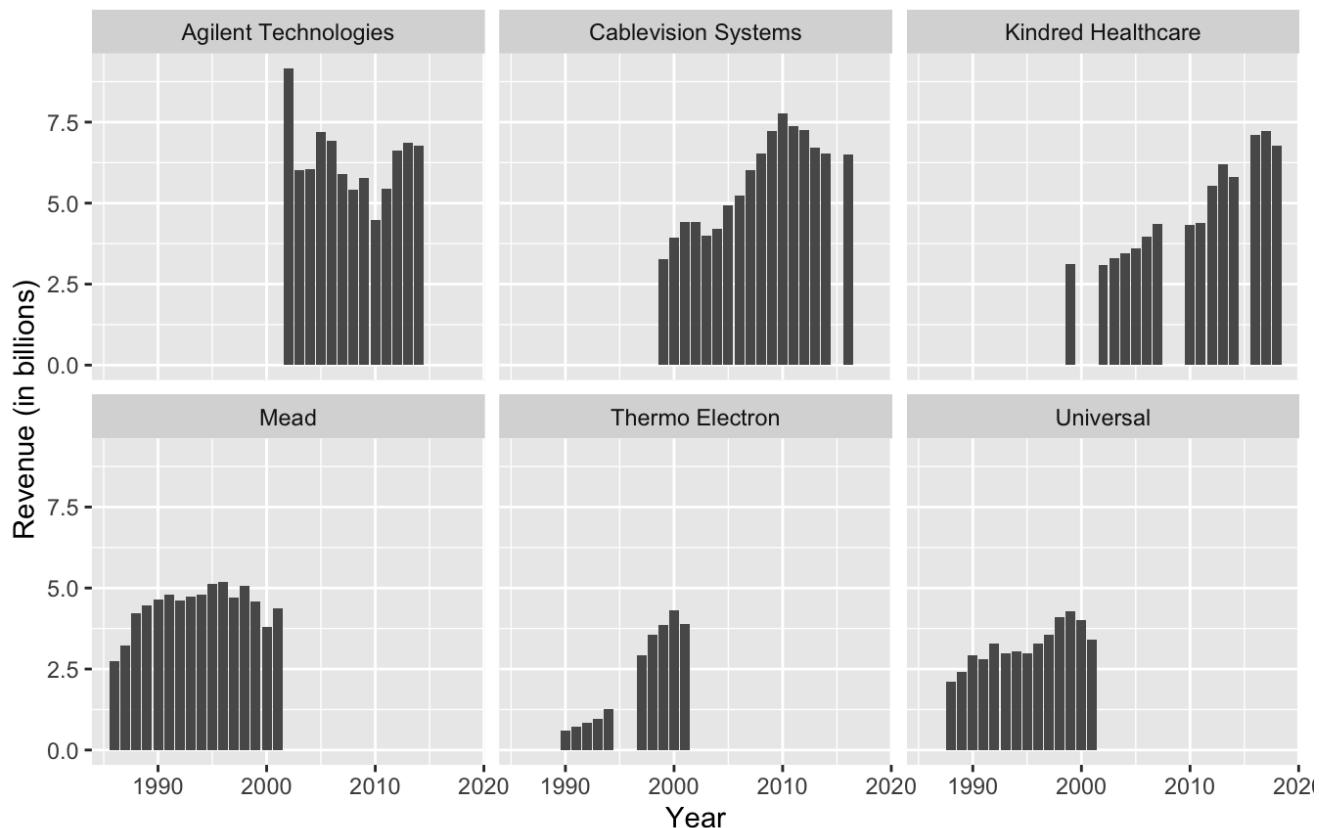
Healthcare Companies



```
# plotting Financials Companies
fortune_joined_v3 %>%
  filter(unsuccessful == "True", Sector == "Healthcare") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Unsuccessful* Companies from 1985 to 2021",
       subtitle = "Financials Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Unsuccessful* Companies from 1985 to 2021

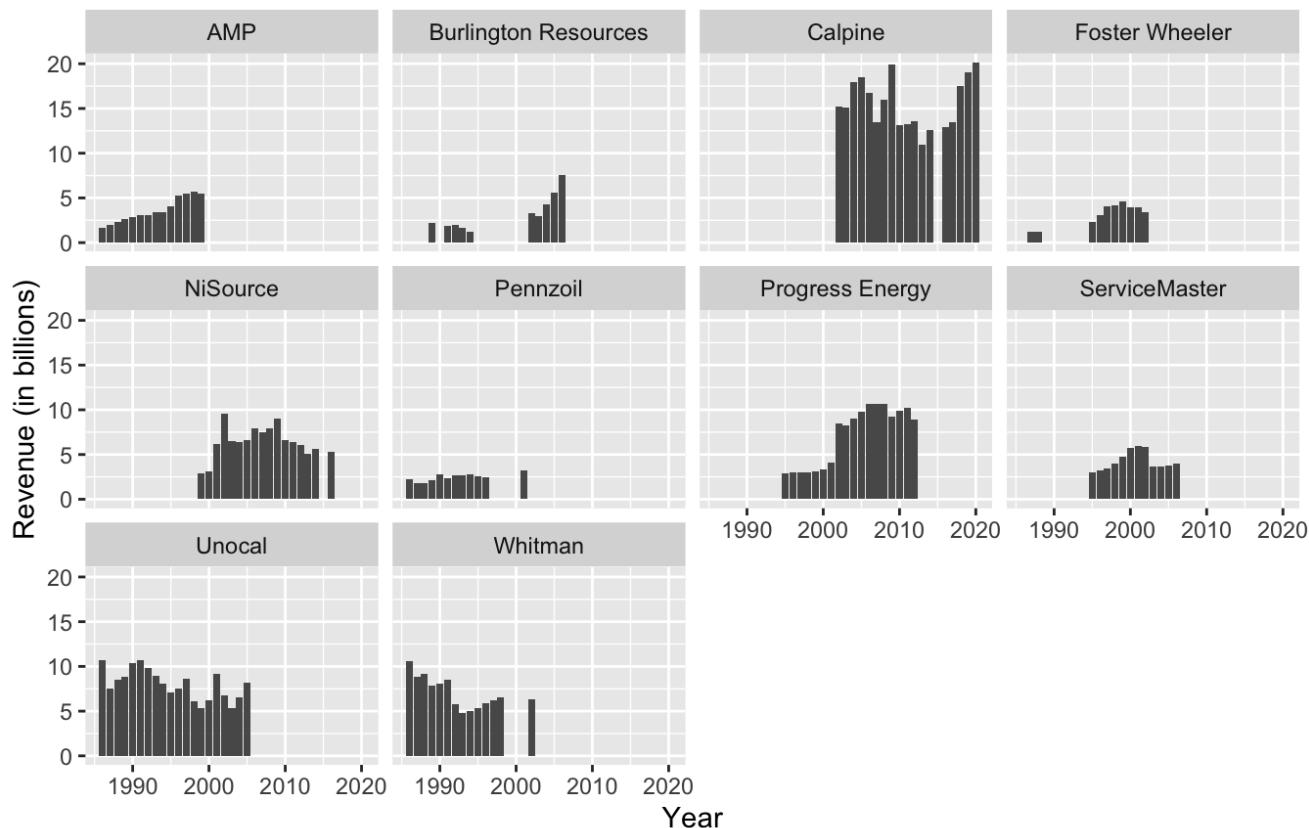
Financials Companies



```
# plotting Energy companies
fortune_joined_v3 %>%
  filter(unsuccessful == "True", Sector == "Energy") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Unsuccessful* Companies from 1985 to 2021",
       subtitle = "Energy Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Unsuccessful* Companies from 1985 to 2021

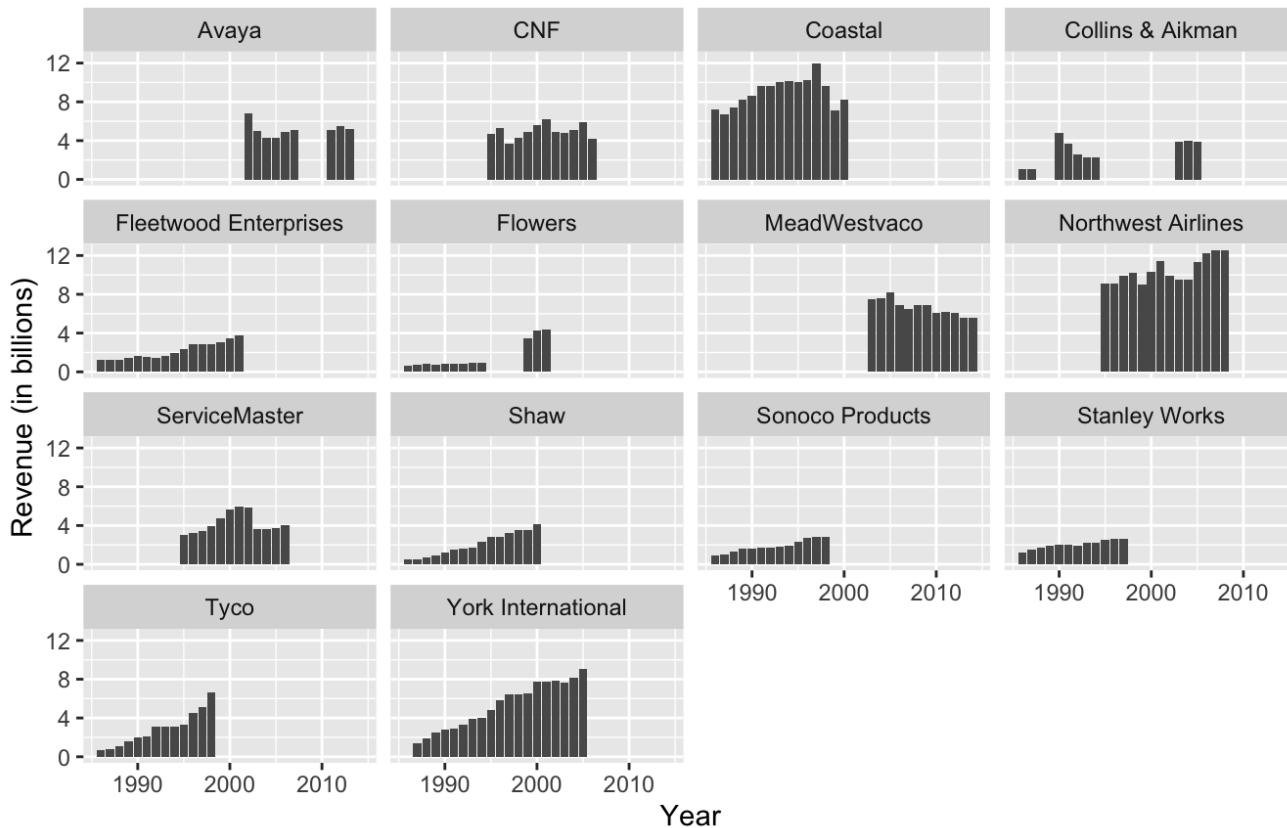
Energy Companies



```
# plotting Industrials companies
fortune_joined_v3 %>%
  filter(unsuccessful == "True", Sector == "Industrials") %>%
  ggplot(aes(Year, Revenue)) +
  geom_col(aes()) +
  facet_wrap(~ Name) +
  labs(title = "Revenue for Unsuccessful* Companies from 1985 to 2021",
       subtitle = "Industrials Companies") +
  ylab("Revenue (in billions)")
```

Revenue for Unsuccessful* Companies from 1985 to 2021

Industrials Companies



Part 5: Technological Influences

In this section, I will start by merging the datasets containing the following information into `fortune_joined_v3`:

1. Supercomputer Power -> `supercomputer_power_flops`
2. Computing Efficiency -> `computing_efficiency`
3. Internet Usage -> `percent_using_internet`
4. Phone Usage -> `landline_cellular_data`
5. Moore's Law -> `moores_law`
6. Computing Cost -> `computer_storage_cost`

Dataset	Source	Description
Technology Adoption	https://ourworldindata.org/grapher/technology-adoption-by-households-in-the-united-states (https://ourworldindata.org/grapher/technology-adoption-by-households-in-the-united-states)	This dataset details the rates of diffusion and adoption of a range of technologies in the United States, measured as the percentage of US households with access or adoption over time.
Supercomputer Power	https://ourworldindata.org/grapher/supercomputer-power-flops (https://ourworldindata.org/grapher/supercomputer-power-flops)	Number of floating-point operations carried out per second by the largest supercomputer in any given year.

Dataset	Source	Description
Computing Efficiency	https://ourworldindata.org/grapher/computing-efficiency (https://ourworldindata.org/grapher/computing-efficiency)	Computer processing efficiency, measured as the number of watts needed per million instructions per second (Watts per MIPS).
Internet Usage	https://data.worldbank.org/indicator/IT.NET.USER.ZS (https://data.worldbank.org/indicator/IT.NET.USER.ZS)	Percent of US population using the internet by year
Phone Usage	https://data.worldbank.org/indicator/IT.CEL.SETS (https://data.worldbank.org/indicator/IT.CEL.SETS)	Data on Cellular and Land line subscriptions in the US
Moore's Law	https://ourworldindata.org/technological-change (https://ourworldindata.org/technological-change)	Real world data on Moore's law, which states the number of transistors on a microchip doubles about every two years, though the cost of computers is halved.
Computing Cost	https://ourworldindata.org/technological-change (https://ourworldindata.org/technological-change)	Historical cost of computer memory and storage, measured in US dollars per megabyte.

Joining the Dataframes

```
# add external dataframes
fortune_joined_v4 <- fortune_joined_v3 %>%
  left_join(., supercomputer_power_flops, by = "Year") %>%
  left_join(., computing_efficiency, by = "Year") %>%
  left_join(., percent_using_internet, by = "Year") %>%
  left_join(., landline_cellular_data, by = "Year") %>%
  left_join(., moores_law, by = "Year") %>%
  left_join(., computer_storage_cost, by = "Year")
```

We now have another data frame called `fortune_joined_v4`, which contains all of the external technological factor data. I will use this data to evaluate the effects of each on different sectors.

Analysis

Hyper-Successful Companies vs. Technological Improvements

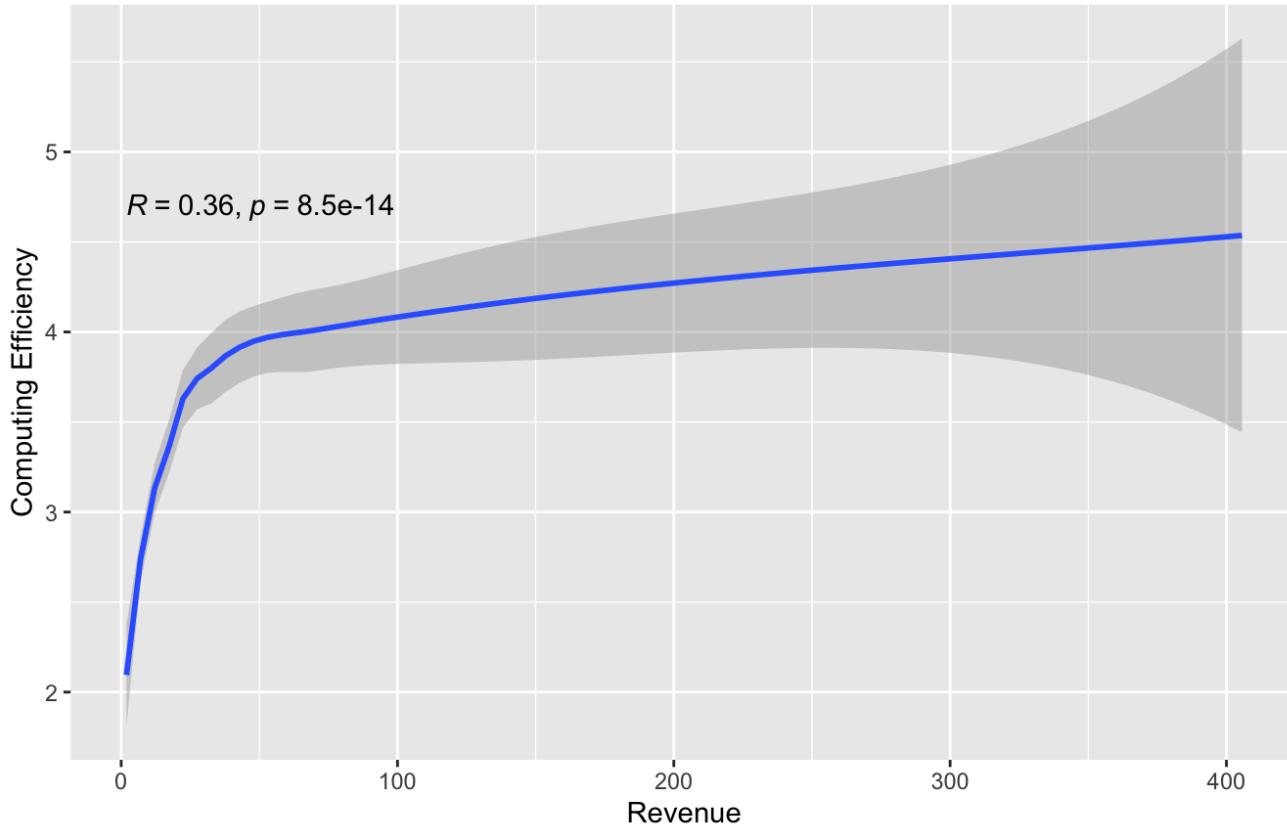
The primary goal of this analysis was to determine what technology factors affect successful companies. Therefore, the relationship between hyper-successful company revenue and technology factors is doubly necessary. The following sections comparing lesser successful companies and unsuccessful companies is aimed at reducing negative effects from confounding variables.

What is the relationship between hyper-successful companies and Technological Improvements?

```
# computing efficiency
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, -log10(`Computing efficiency`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Efficiency Time Series",
       subtitle = "Computing Efficiency vs. Revenue from Hyper-Successful Companies") +
  ylab("Computing Efficiency")
```

Computing Efficiency Time Series

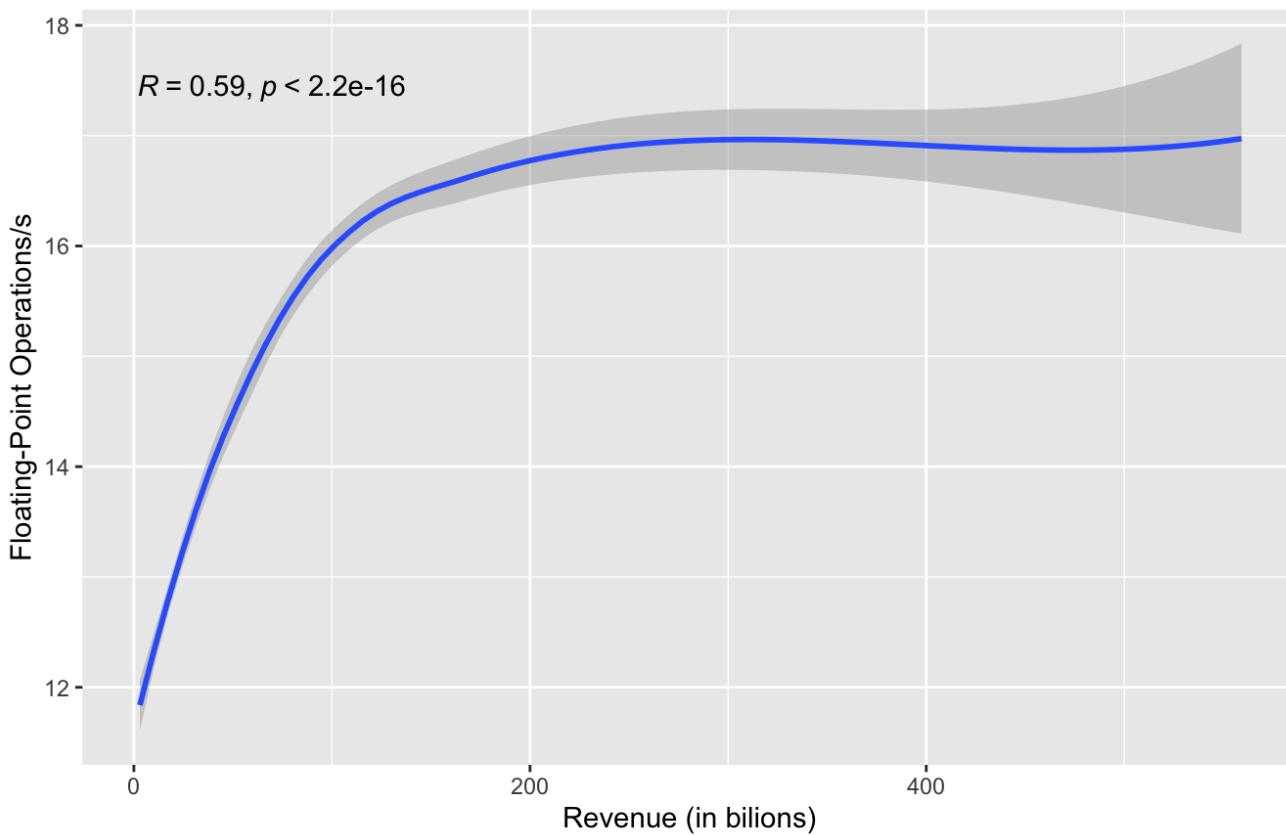
Computing Efficiency vs. Revenue from Hyper-Successful Companies



```
# supercomputer power
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, log10(`Floating-Point Operations per Second`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Supercomputer Power Time Series",
       subtitle = "Floating-Point Operations/s vs. Revenue from Hyper-Successful Companies") +
  ylab("Floating-Point Operations/s") +
  xlab("Revenue (in billions)")
```

Supercomputer Power Time Series

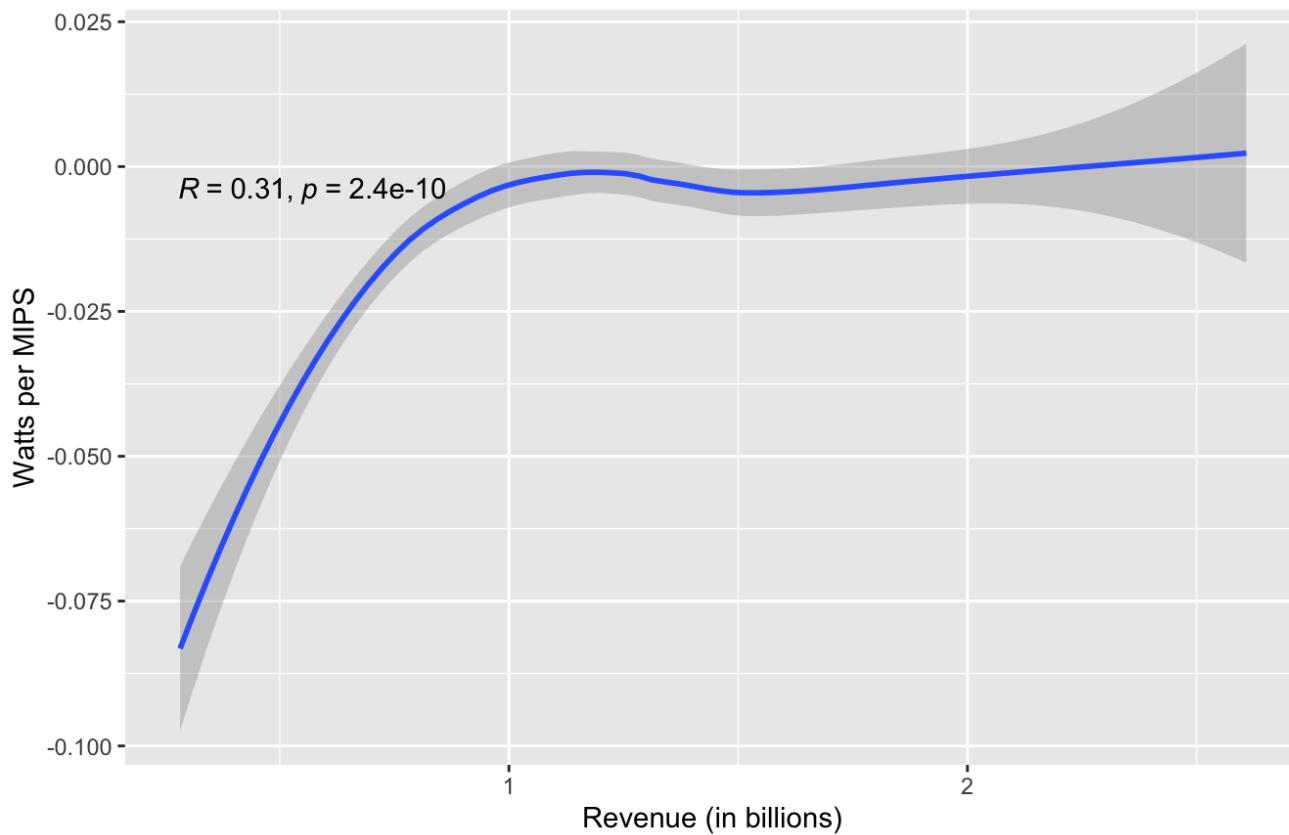
Floating-Point Operations/s vs. Revenue from Hyper-Successful Companies



```
# computing efficiency
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(log10(Revenue), -`Computing efficiency`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Efficiency",
       subtitle = "Watts per MIPS vs. Revenue from Hyper-Successful Companies") +
  ylab("Watts per MIPS") +
  xlab("Revenue (in billions)")
```

Computing Efficiency

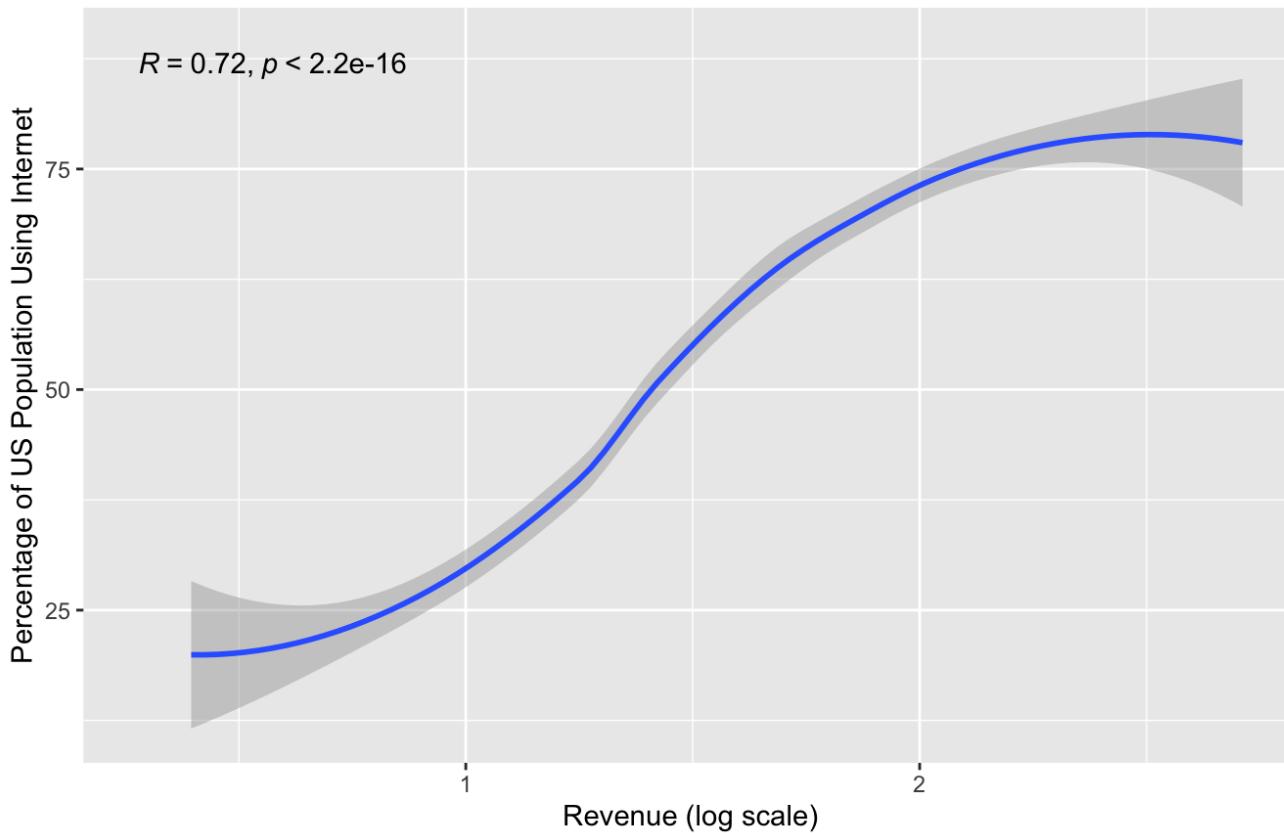
Watts per MIPS vs. Revenue from Hyper-Successful Companies



```
# percent of US using the internet
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(log10(Revenue), `Individuals using the Internet (% of population)`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Internet Usage Time Series",
       subtitle = "Internet Usage vs. Revenue from Hyper-Successful Companies") +
  ylab("Percentage of US Population Using Internet") +
  xlab("Revenue (log scale)")
```

Internet Usage Time Series

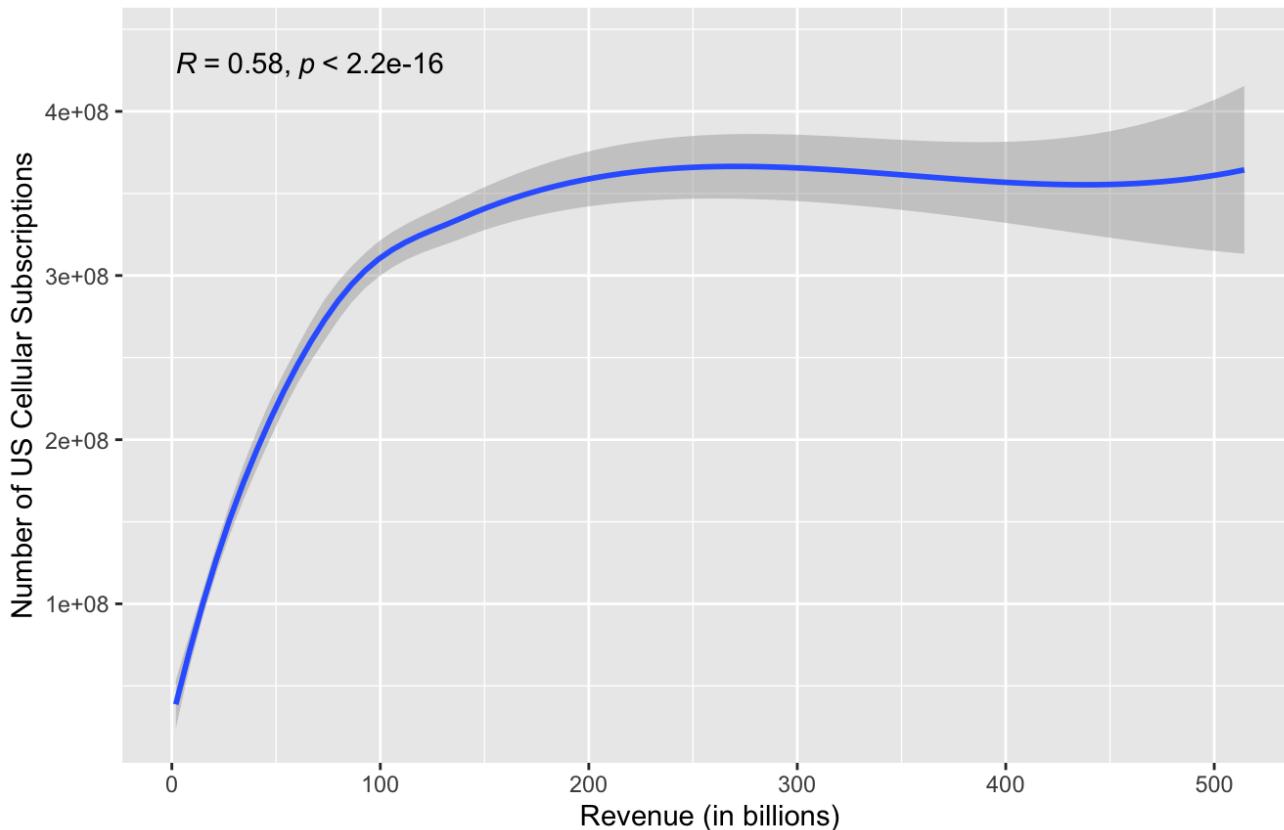
Internet Usage vs. Revenue from Hyper-Successful Companies



```
# Cellular Subscriptions
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, `Cellular Subscriptions`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "US Cellular Subscriptions Time Series",
       subtitle = "Number of Cellular Subscriptions vs. Revenue from Hyper-Successful Companies") +
  ylab("Number of US Cellular Subscriptions") +
  xlab("Revenue (in billions)")
```

US Cellular Subscriptions Time Series

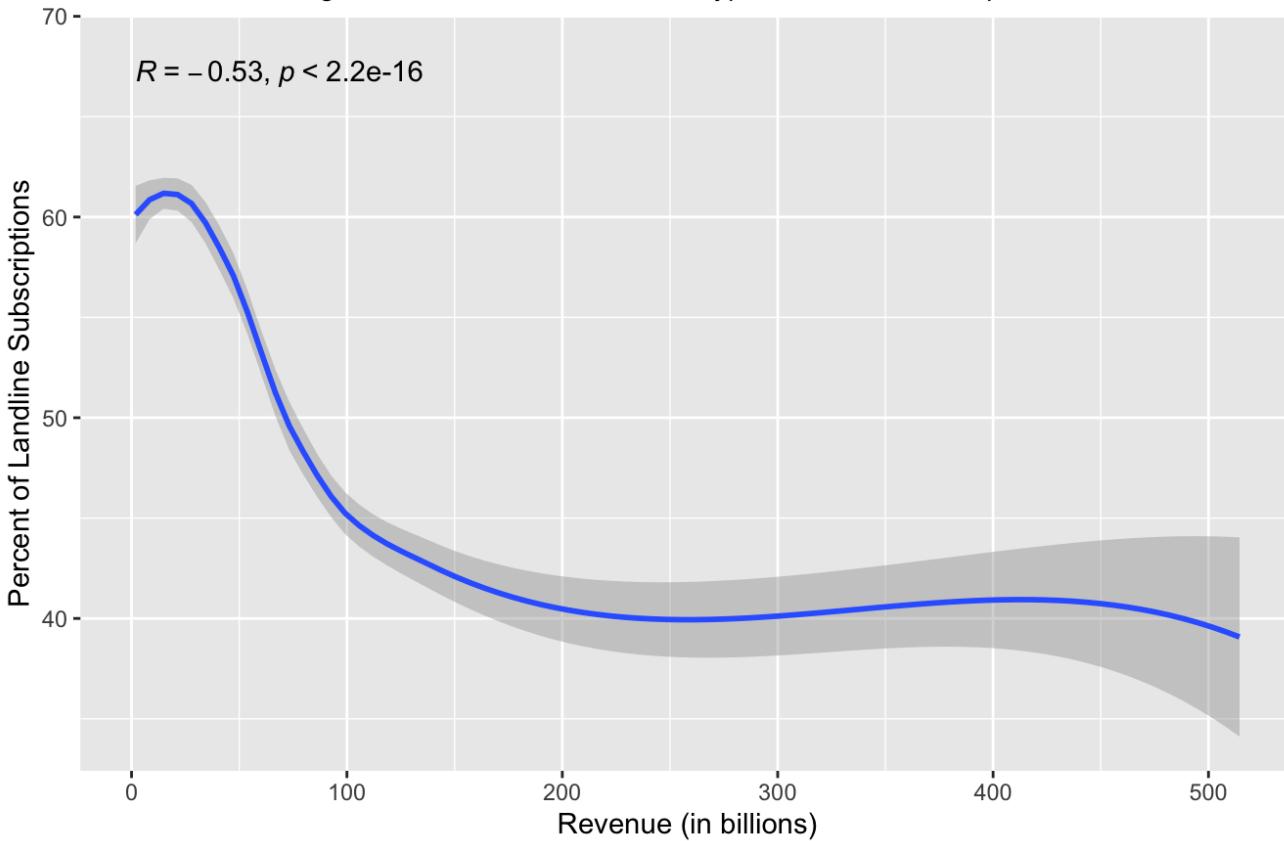
Number of Cellular Subscriptions vs. Revenue from Hyper-Successful Companies



```
# landline subscriptions (per 100 people)
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, `Landline Subscriptions (per 100 people)`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "US Landline Subscriptions Time Series",
       subtitle = "Percent of US using Landline vs. Revenue from Hyper-Successful Companies") +
  ylab("Percent of Landline Subscriptions") +
  xlab("Revenue (in billions)")
```

US Landline Subscriptions Time Series

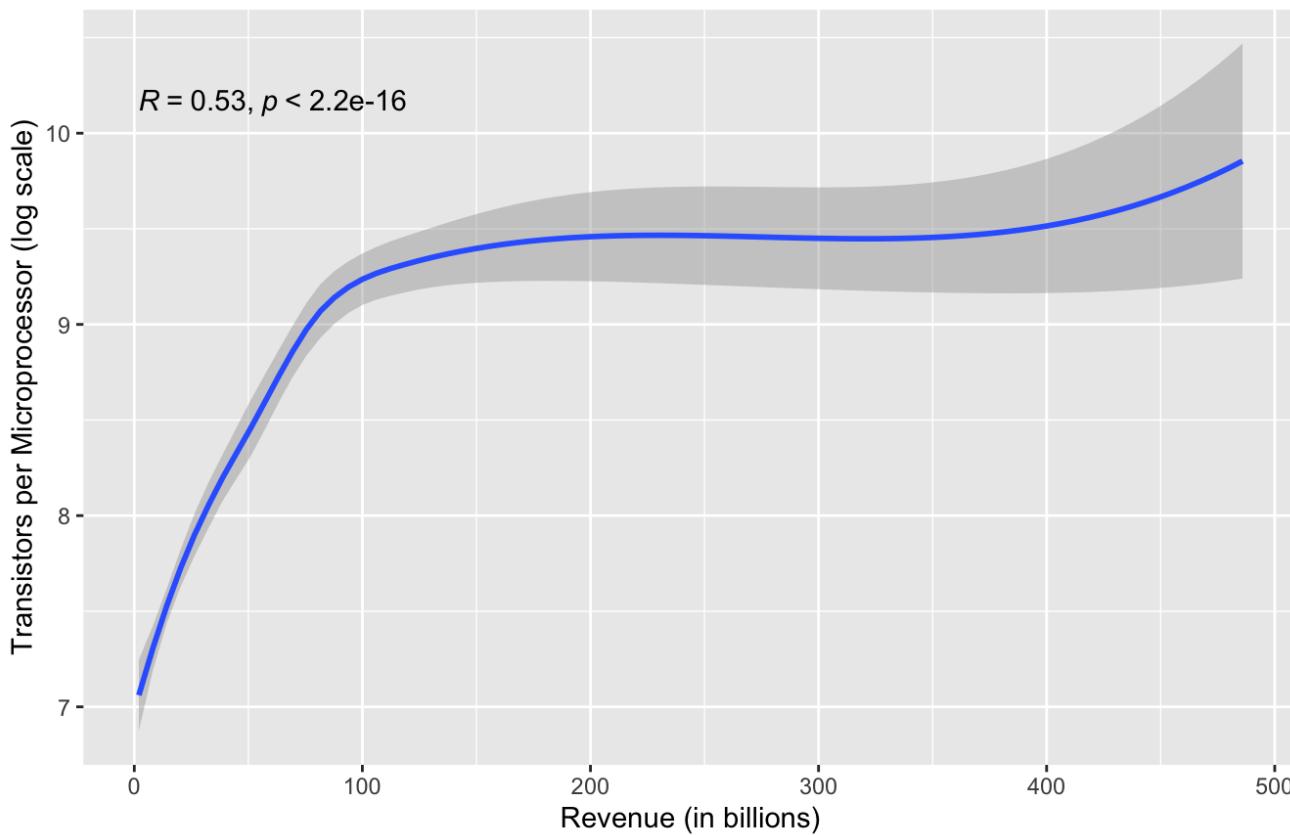
Percent of US using Landline vs. Revenue from Hyper-Successful Companies



```
# moores law
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, log10(`Transistors per microprocessor`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Mooore's Law Time Series",
       subtitle = "Transistors per Microprocessor vs. Revenue from Hyper-Successful Companies") +
  ylab("Transistors per Microprocessor (log scale)") +
  xlab("Revenue (in billions)")
```

Mooore's Law Time Series

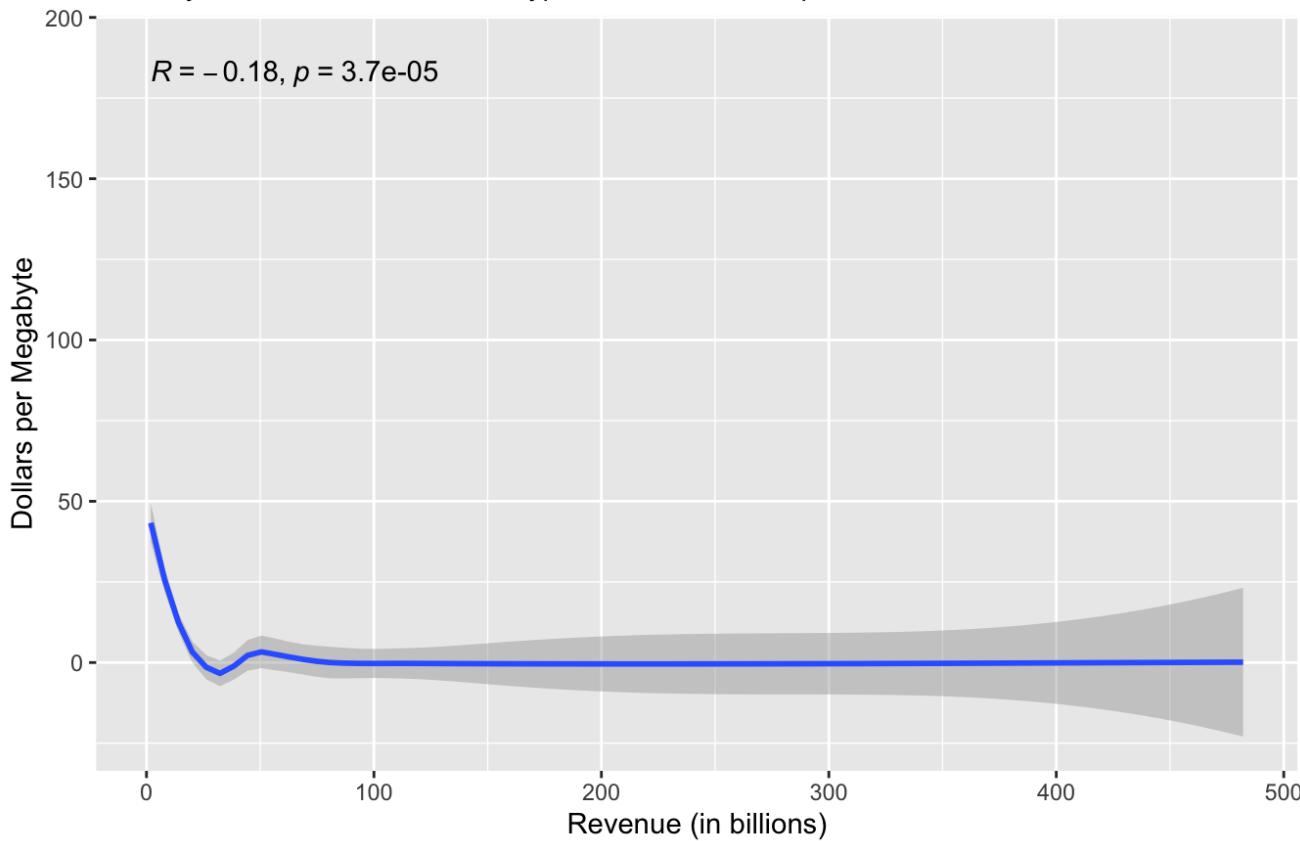
Transistors per Microprocessor vs. Revenue from Hyper-Successful Companies



```
# computing cost (memory)
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, memory)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Cost Time Series",
       subtitle = "Memory Cost vs. Revenue from Hyper-Successful Companies") +
  ylab("Dollars per Megabyte") +
  xlab("Revenue (in billions)")
```

Computing Cost Time Series

Memory Cost vs. Revenue from Hyper-Successful Companies

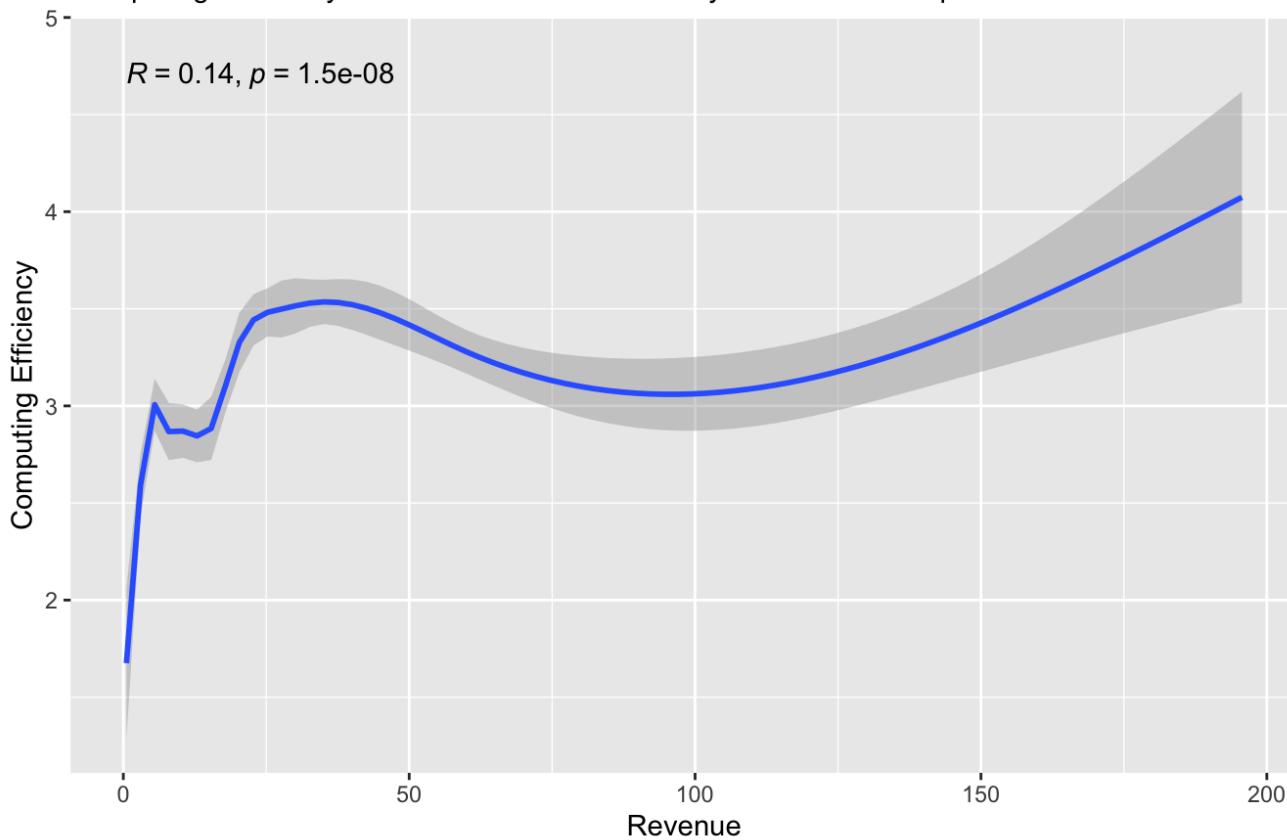


Moderately-Successful Companies vs. Technological Improvements

```
# computing efficiency
fortune_joined_v4 %>%
  filter(moderate_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, -log10(`Computing efficiency`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Efficiency Time Series",
       subtitle = "Computing Efficiency vs. Revenue from Moderately-Successful Companies") +
  ylab("Computing Efficiency")
```

Computing Efficiency Time Series

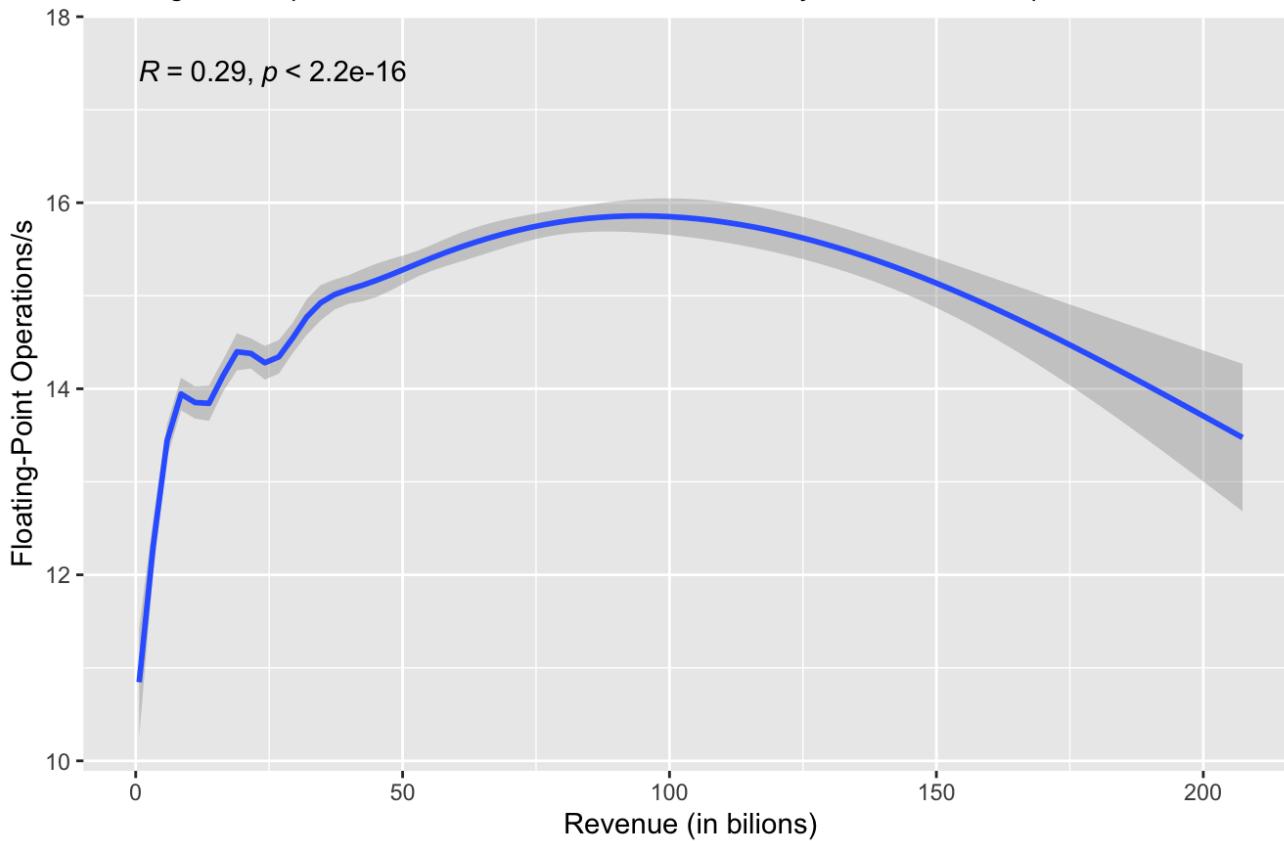
Computing Efficiency vs. Revenue from Moderately-Successful Companies



```
# supercomputer power
fortune_joined_v4 %>%
  filter(moderate_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, log10(`Floating-Point Operations per Second`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Supercomputer Power Time Series",
       subtitle = "Floating-Point Operations/s vs. Revenue from Moderately-Successful Companies") +
  ylab("Floating-Point Operations/s") +
  xlab("Revenue (in billions)")
```

Supercomputer Power Time Series

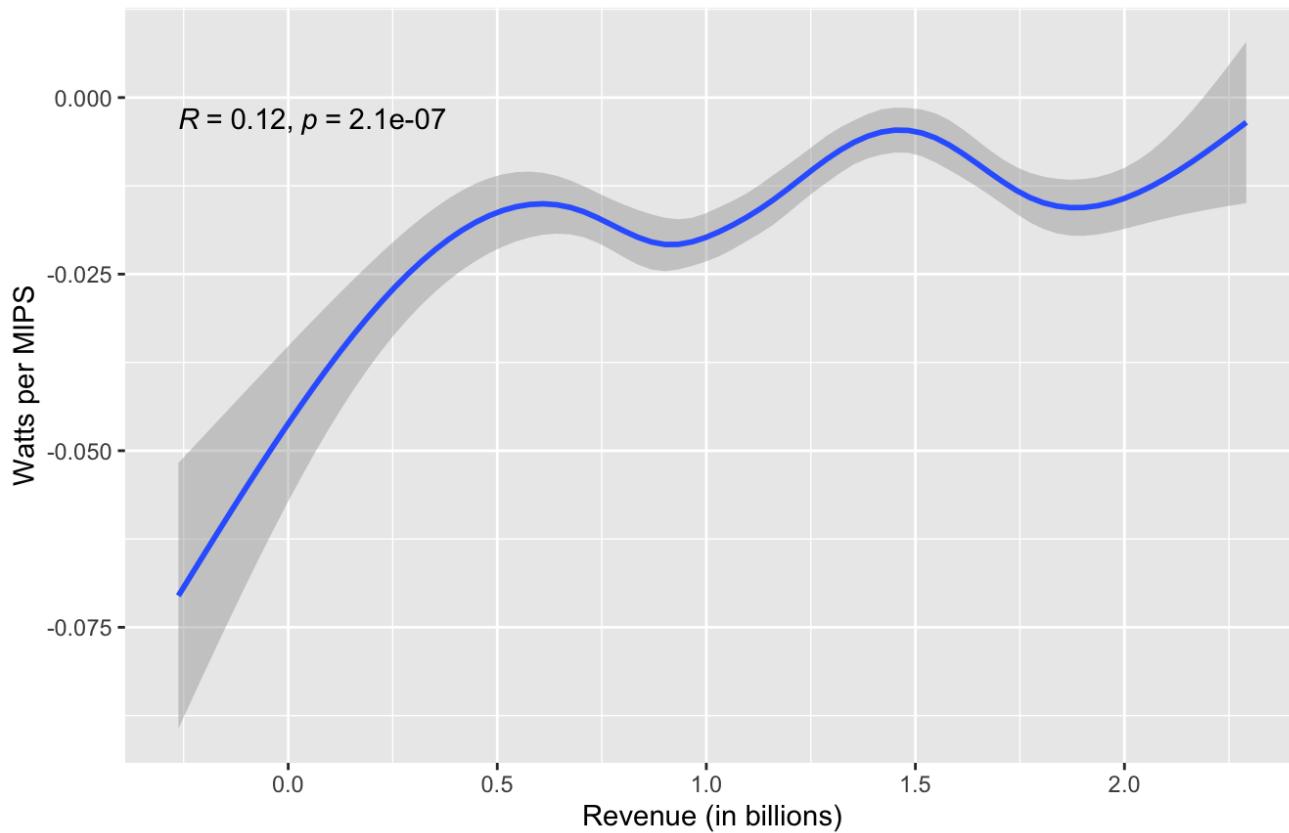
Floating-Point Operations/s vs. Revenue from Moderately-Successful Companies



```
# computing efficiency
fortune_joined_v4 %>%
  filter(moderate_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(log10(Revenue), -`Computing efficiency`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Efficiency",
       subtitle = "Watts per MIPS vs. Revenue from Moderately-Successful Companies") +
  ylab("Watts per MIPS") +
  xlab("Revenue (in billions)")
```

Computing Efficiency

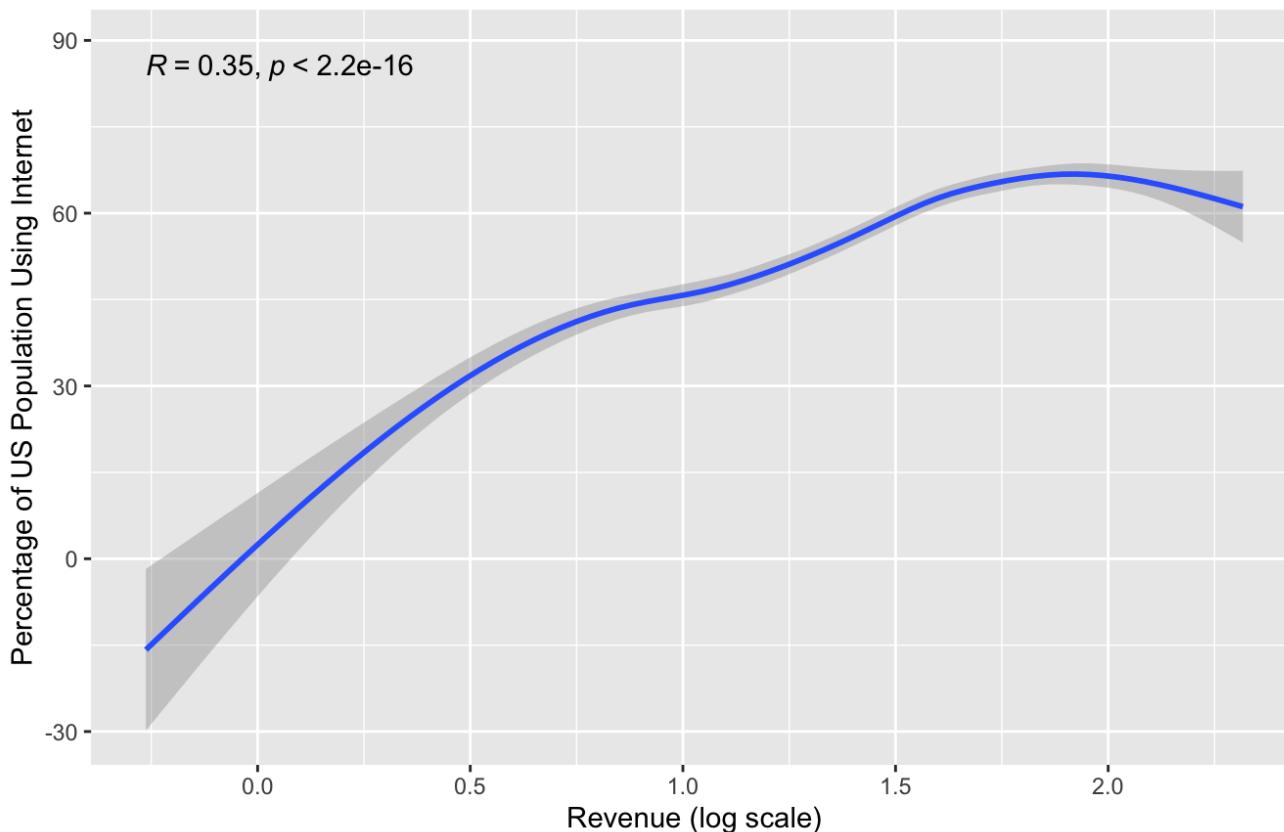
Watts per MIPS vs. Revenue from Moderately-Successful Companies



```
# percent of US using the internet
fortune_joined_v4 %>%
  filter(moderate_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(log10(Revenue), `Individuals using the Internet (% of population)`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Internet Usage Time Series",
       subtitle = "Internet Usage vs. Revenue from Moderately-Successful Companies") +
  ylab("Percentage of US Population Using Internet") +
  xlab("Revenue (log scale)")
```

Internet Usage Time Series

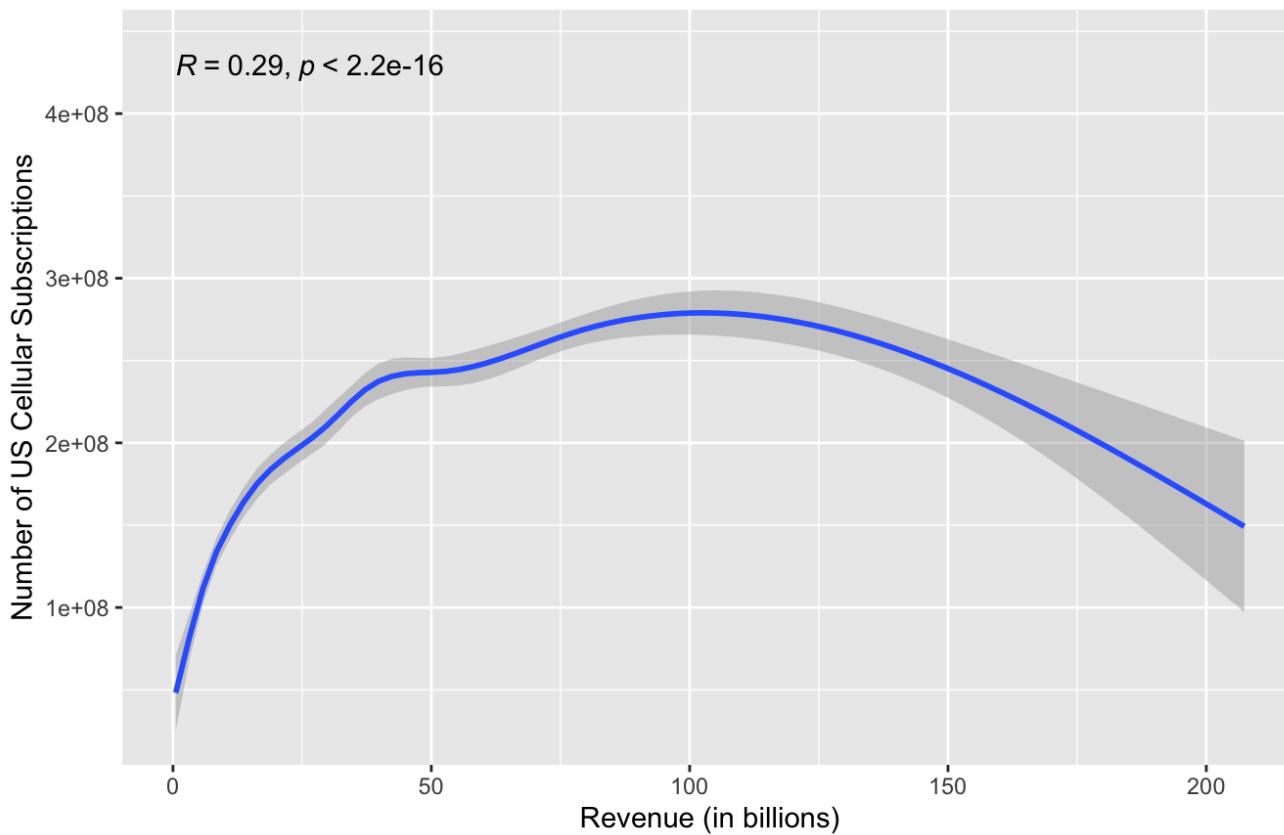
Internet Usage vs. Revenue from Moderately-Successful Companies



```
# Cellular Subscriptions
fortune_joined_v4 %>%
  filter(moderate_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, `Cellular Subscriptions`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "US Cellular Subscriptions Time Series",
       subtitle = "Number of Cellular Subscriptions vs. Revenue from Moderately-Successful
Companies") +
  ylab("Number of US Cellular Subscriptions") +
  xlab("Revenue (in billions)")
```

US Cellular Subscriptions Time Series

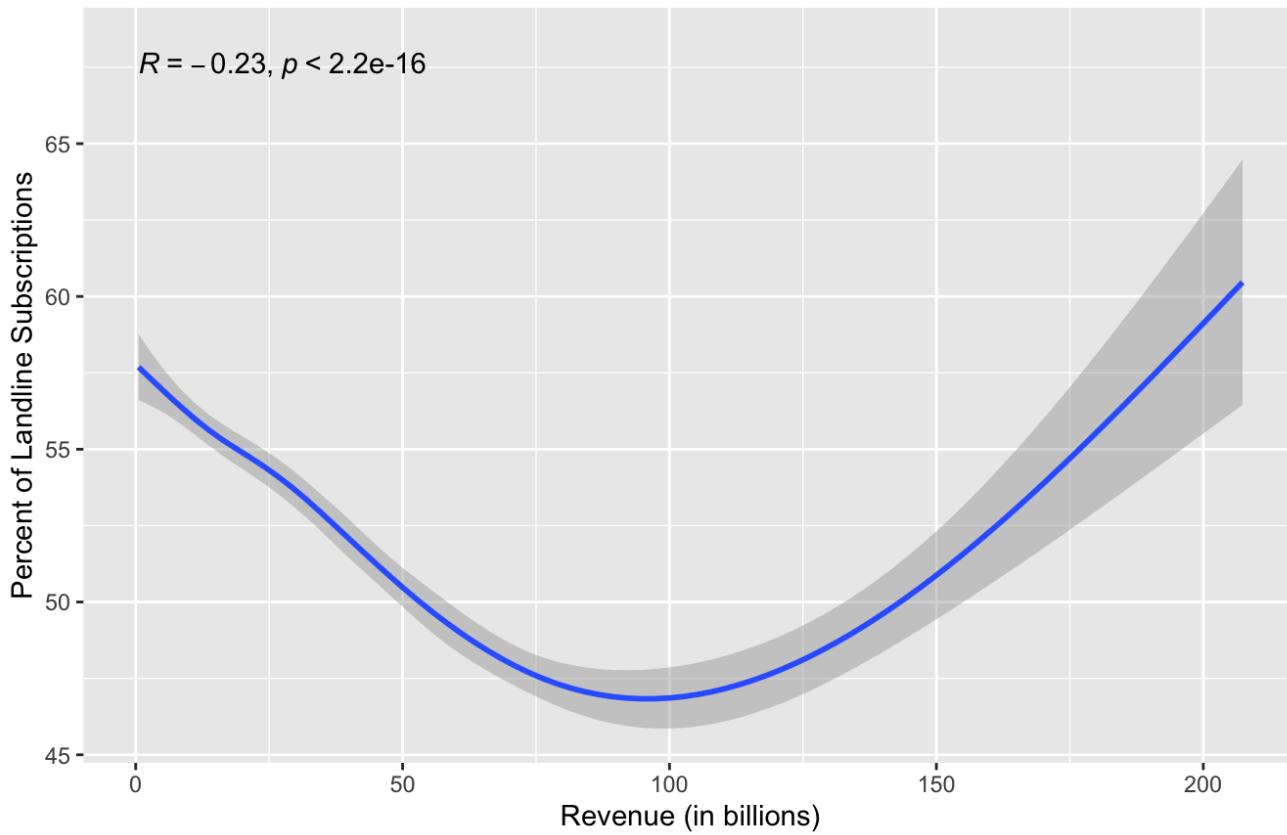
Number of Cellular Subscriptions vs. Revenue from Moderately-Successful Companies



```
# landline subscriptions (per 100 people)
fortune_joined_v4 %>%
  filter(moderate_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, `Landline Subscriptions (per 100 people)`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "US Landline Subscriptions Time Series",
       subtitle = "Percent of US using Landline vs. Revenue from Moderately-Successful Companies") +
  ylab("Percent of Landline Subscriptions") +
  xlab("Revenue (in billions)")
```

US Landline Subscriptions Time Series

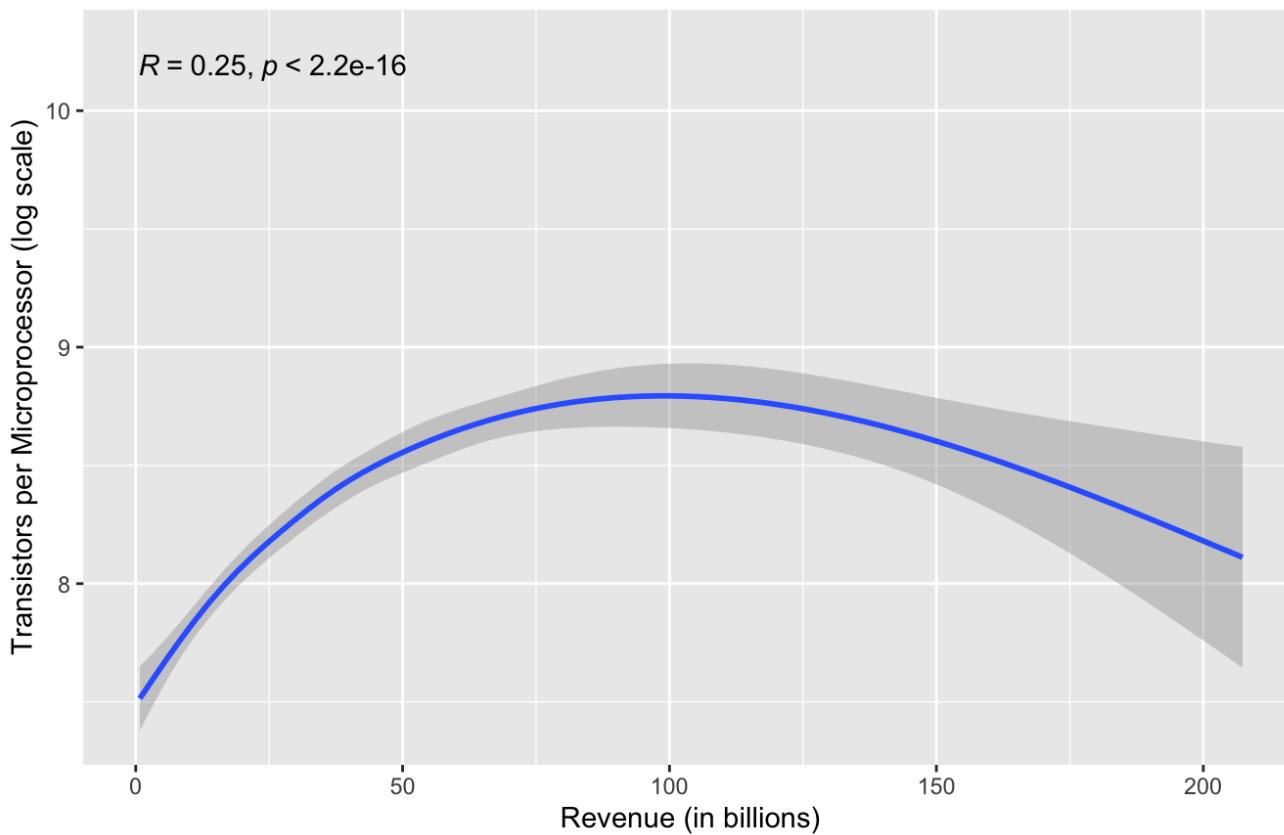
Percent of US using Landline vs. Revenue from Moderately-Successful Companies



```
# moores law
fortune_joined_v4 %>%
  filter(moderate_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, log10(`Transistors per microprocessor`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Mooore's Law Time Series",
       subtitle = "Transistors per Microprocessor vs. Revenue from Moderately-Successful Companies") +
  ylab("Transistors per Microprocessor (log scale)") +
  xlab("Revenue (in billions)")
```

Mooore's Law Time Series

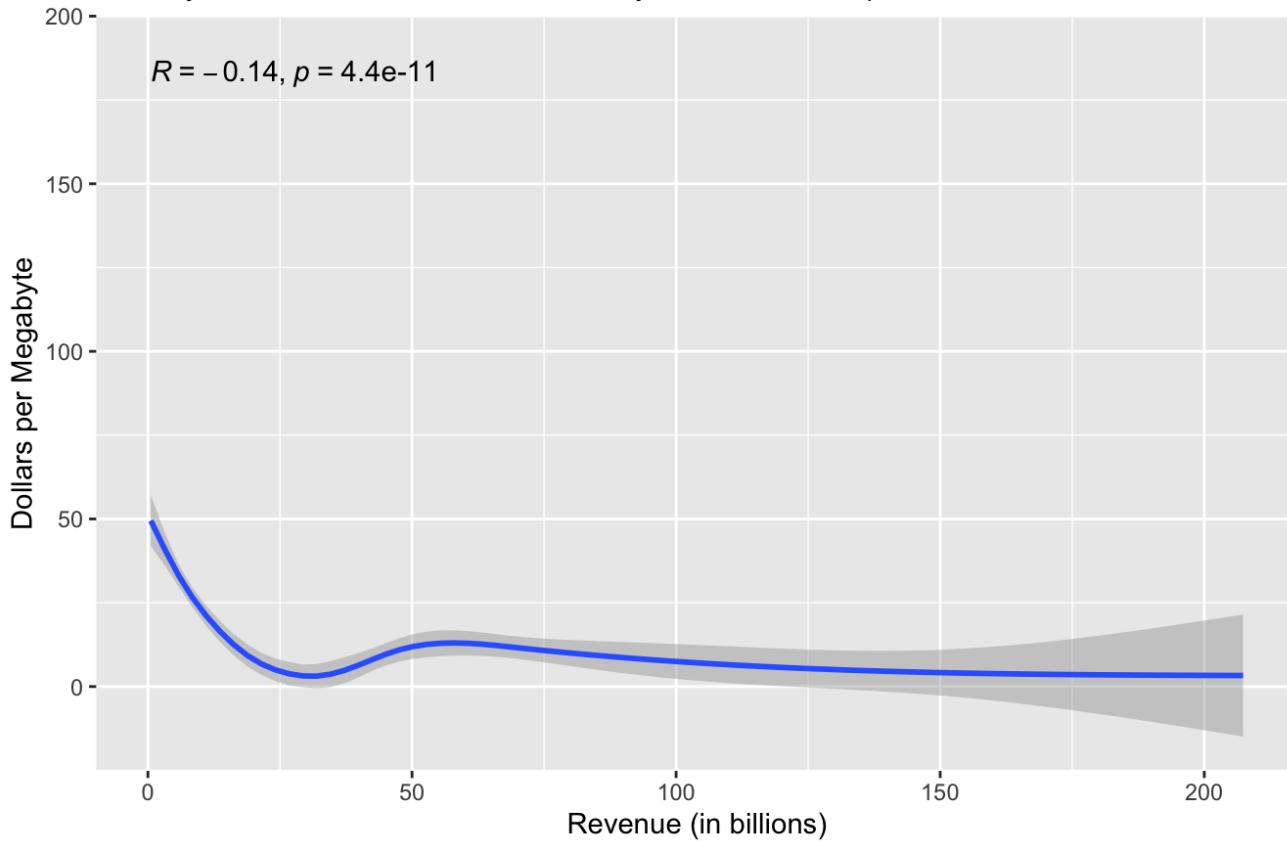
Transistors per Microprocessor vs. Revenue from Moderately-Successful Companies



```
# computing cost (memory)
fortune_joined_v4 %>%
  filter(moderate_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, memory)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Cost Time Series",
       subtitle = "Memory Cost vs. Revenue from Moderately-Successful Companies") +
  ylab("Dollars per Megabyte") +
  xlab("Revenue (in billions)")
```

Computing Cost Time Series

Memory Cost vs. Revenue from Moderately-Successful Companies

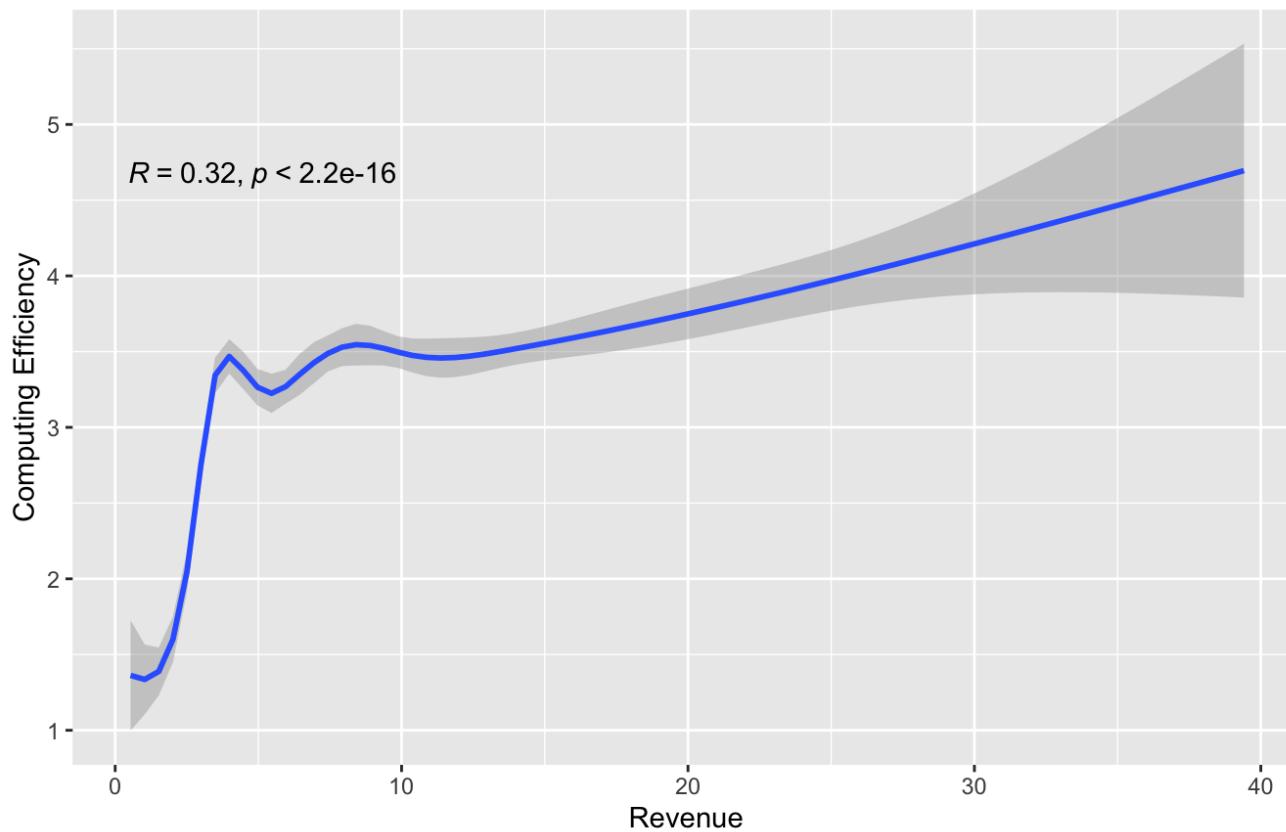


Low-Growth Companies vs. Technological Improvements

```
# computing efficiency
fortune_joined_v4 %>%
  filter(low_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, -log10(`Computing efficiency`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Efficiency Time Series",
       subtitle = "Computing Efficiency vs. Revenue from Low-Growth Companies") +
  ylab("Computing Efficiency")
```

Computing Efficiency Time Series

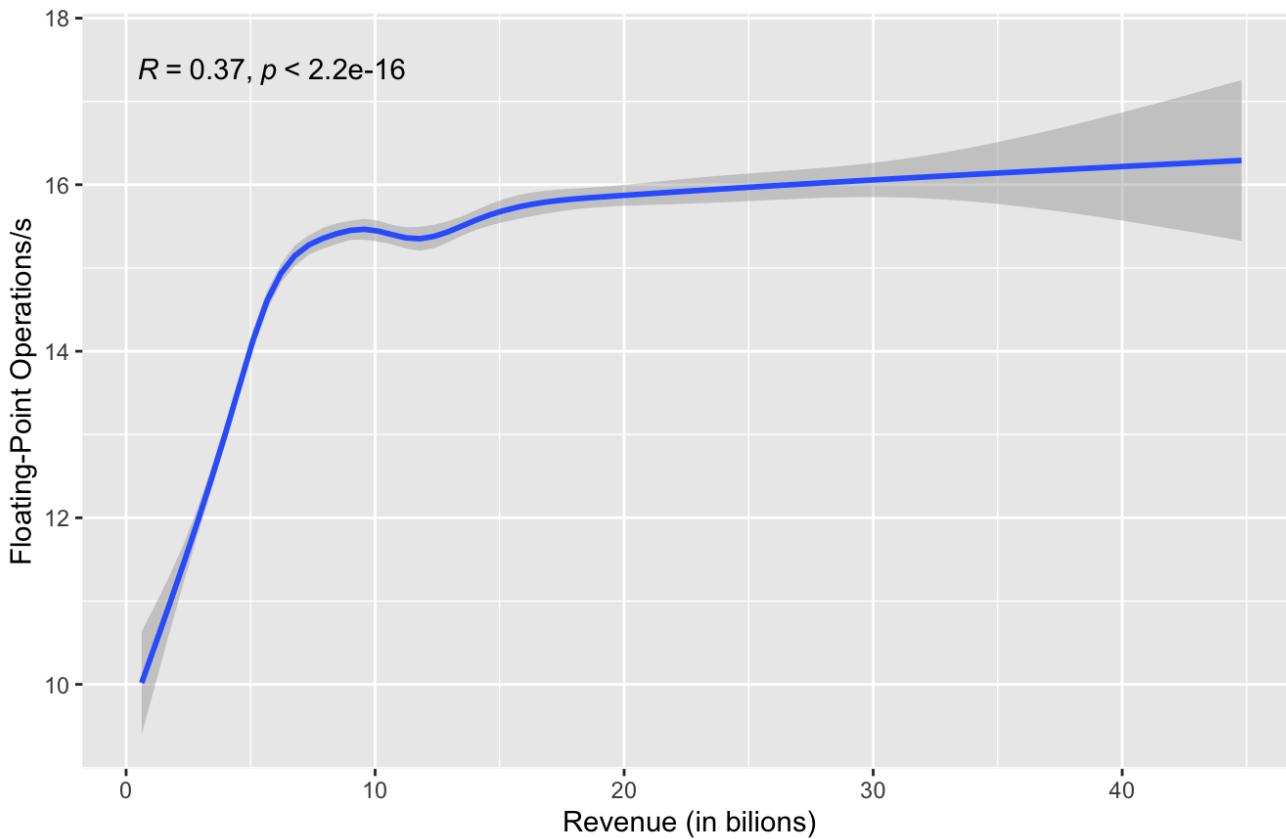
Computing Efficiency vs. Revenue from Low-Growth Companies



```
# supercomputer power
fortune_joined_v4 %>%
  filter(low_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, log10(`Floating-Point Operations per Second`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Supercomputer Power Time Series",
       subtitle = "Floating-Point Operations/s vs. Revenue from Low-Growth Companies") +
  ylab("Floating-Point Operations/s") +
  xlab("Revenue (in billions)")
```

Supercomputer Power Time Series

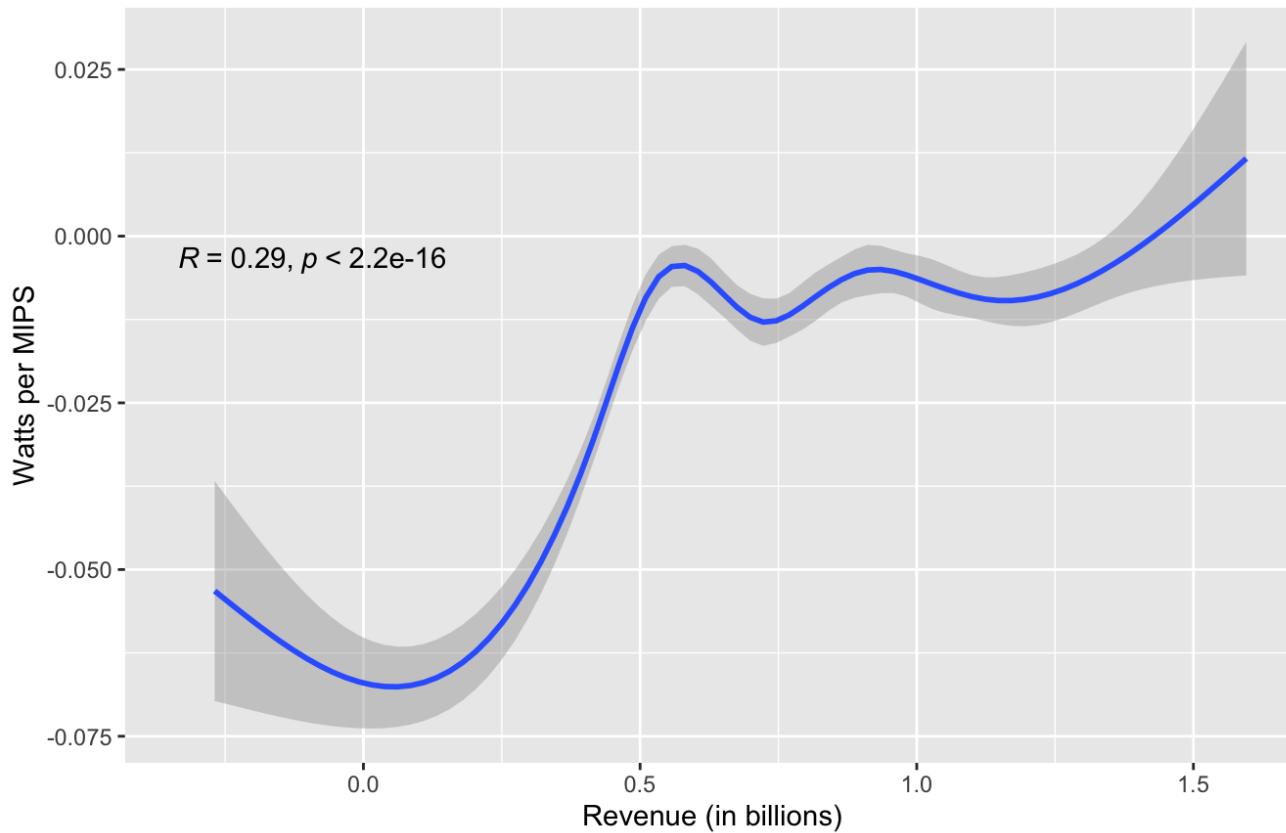
Floating-Point Operations/s vs. Revenue from Low-Growth Companies



```
# computing efficiency
fortune_joined_v4 %>%
  filter(low_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(log10(Revenue), -`Computing efficiency`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Efficiency",
       subtitle = "Watts per MIPS vs. Revenue from Low-Growth Companies") +
  ylab("Watts per MIPS") +
  xlab("Revenue (in billions)")
```

Computing Efficiency

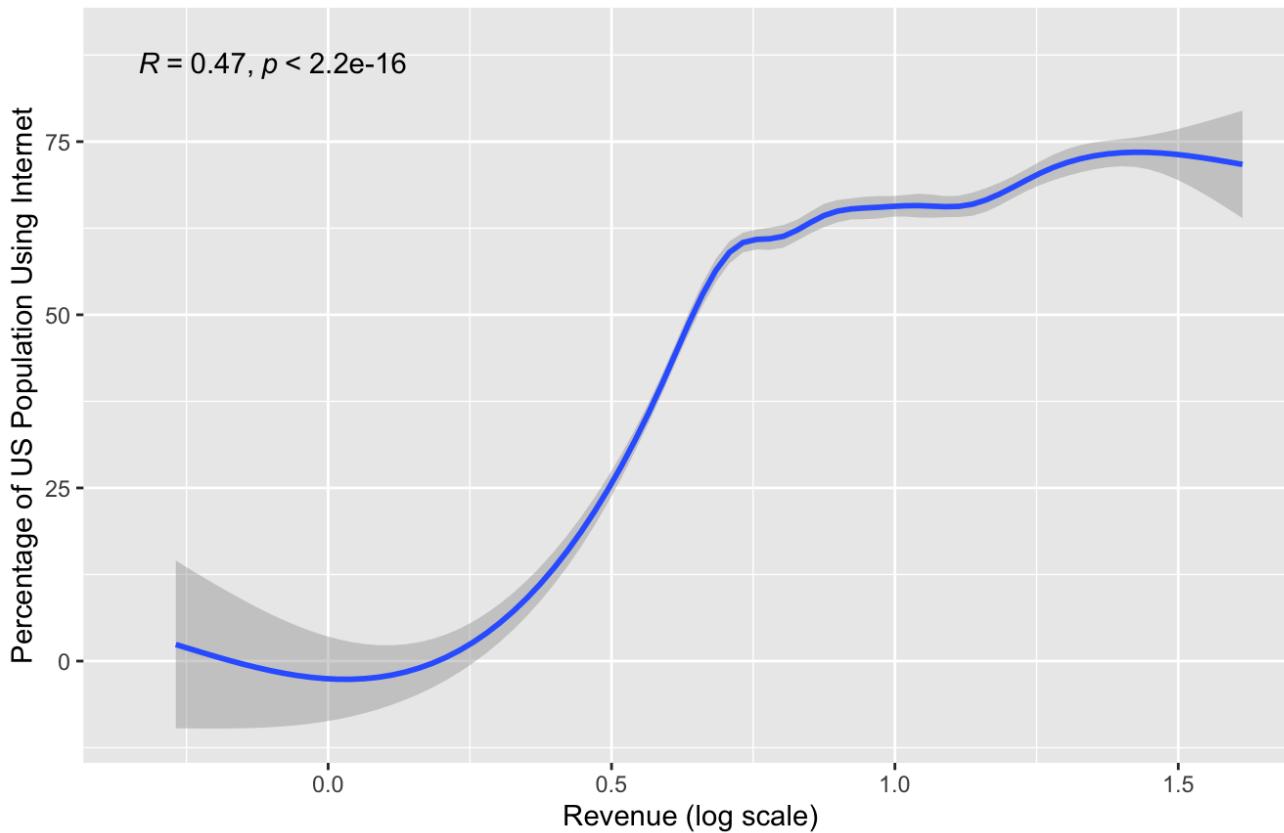
Watts per MIPS vs. Revenue from Low-Growth Companies



```
# percent of US using the internet
fortune_joined_v4 %>%
  filter(low_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(log10(Revenue), `Individuals using the Internet (% of population)`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Internet Usage Time Series",
       subtitle = "Internet Usage vs. Revenue from Low-Growth Companies") +
  ylab("Percentage of US Population Using Internet") +
  xlab("Revenue (log scale)")
```

Internet Usage Time Series

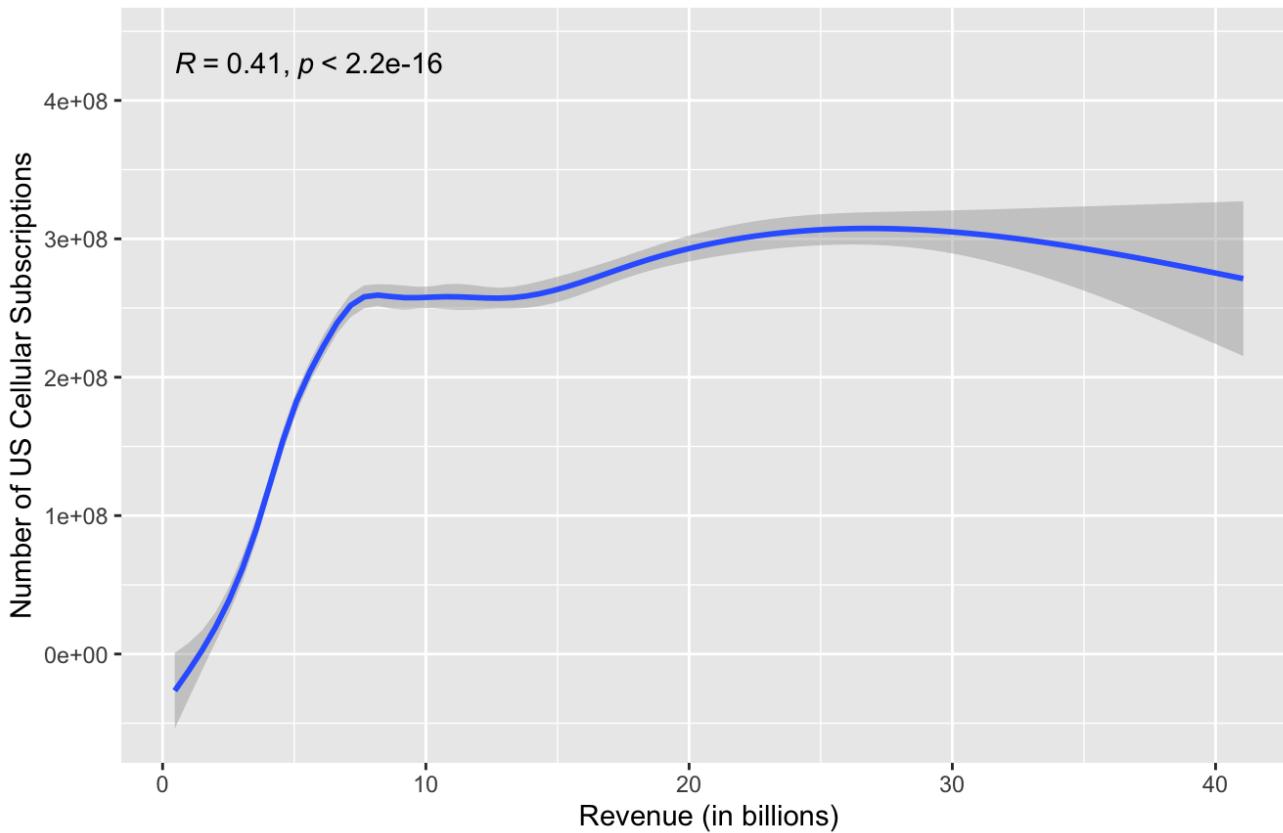
Internet Usage vs. Revenue from Low-Growth Companies



```
# Cellular Subscriptions
fortune_joined_v4 %>%
  filter(low_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, `Cellular Subscriptions`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "US Cellular Subscriptions Time Series",
       subtitle = "Number of Cellular Subscriptions vs. Revenue from Low-Growth Companies"
  ) +
  ylab("Number of US Cellular Subscriptions") +
  xlab("Revenue (in billions)")
```

US Cellular Subscriptions Time Series

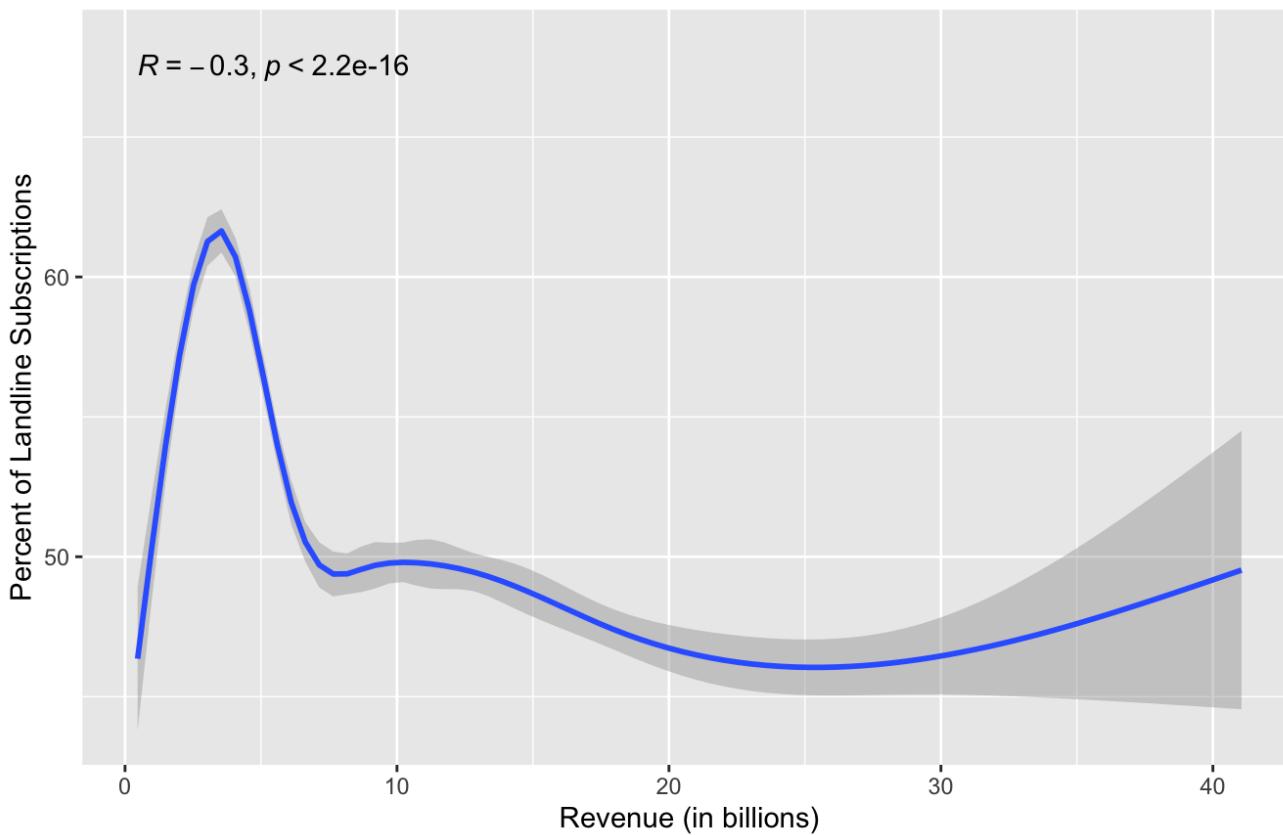
Number of Cellular Subscriptions vs. Revenue from Low-Growth Companies



```
# landline subscriptions (per 100 people)
fortune_joined_v4 %>%
  filter(low_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, `Landline Subscriptions (per 100 people)`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "US Landline Subscriptions Time Series",
       subtitle = "Percent of US using Landline vs. Revenue from Low-Growth Companies") +
  ylab("Percent of Landline Subscriptions") +
  xlab("Revenue (in billions)")
```

US Landline Subscriptions Time Series

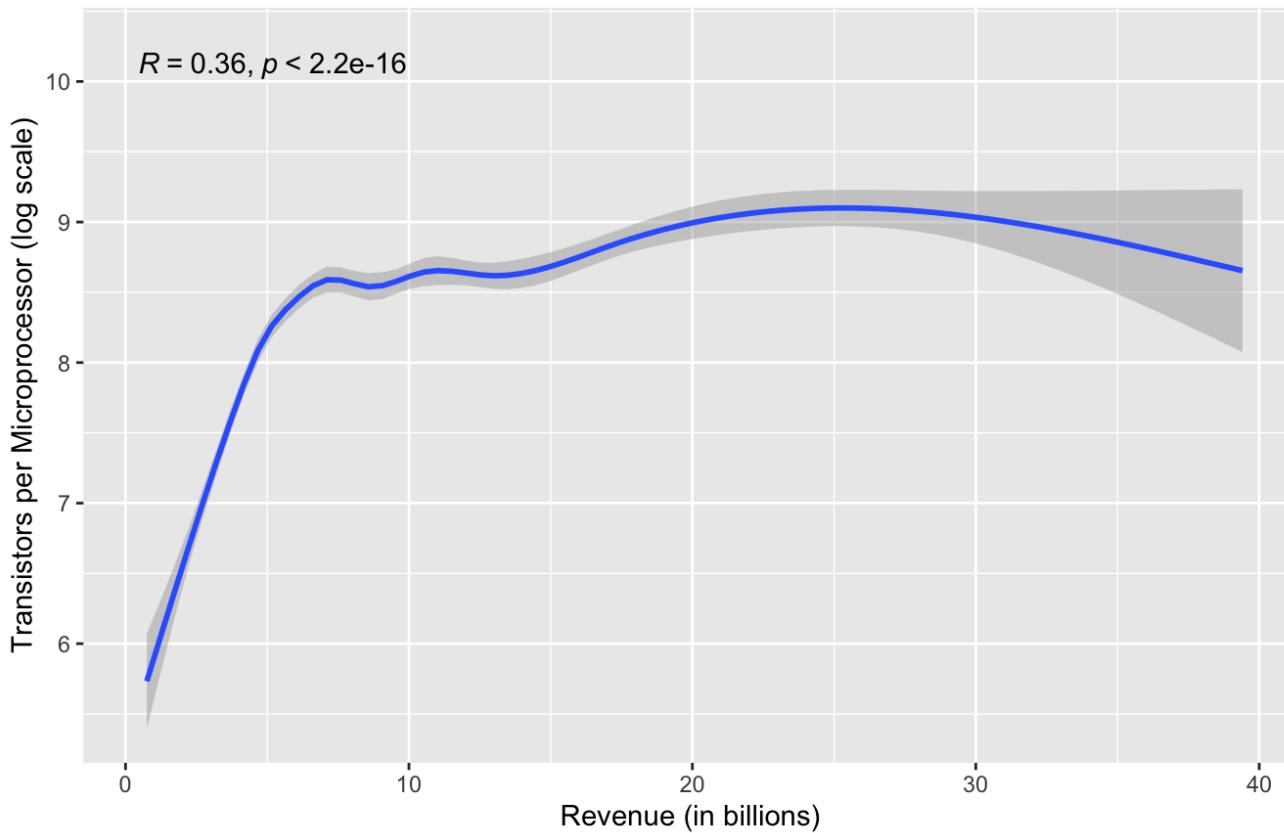
Percent of US using Landline vs. Revenue from Low-Growth Companies



```
# moores law
fortune_joined_v4 %>%
  filter(low_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, log10(`Transistors per microprocessor`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Mooore's Law Time Series",
       subtitle = "Transistors per Microprocessor vs. Revenue from Low-Growth Companies") +
  ylab("Transistors per Microprocessor (log scale)") +
  xlab("Revenue (in billions)")
```

Mooore's Law Time Series

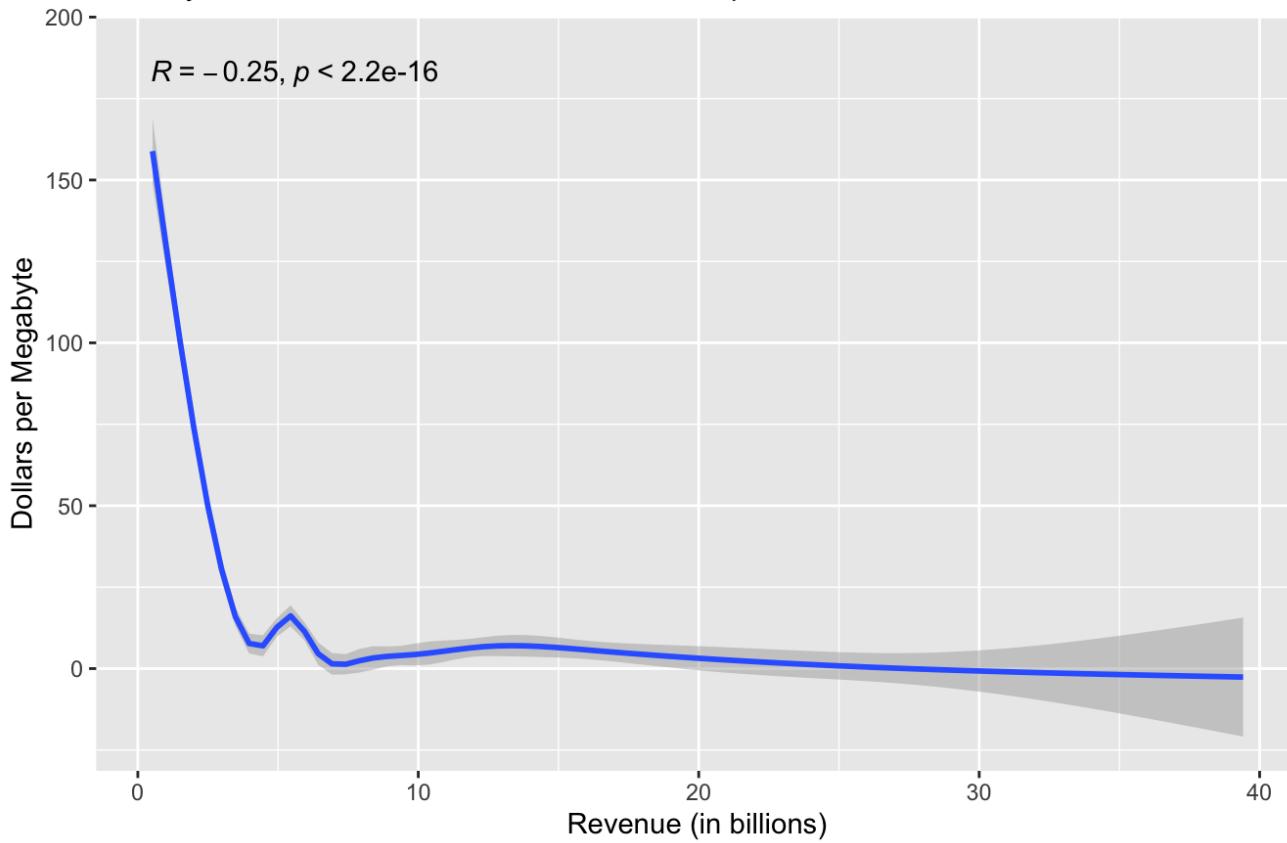
Transistors per Microprocessor vs. Revenue from Low-Growth Companies



```
# computing cost (memory)
fortune_joined_v4 %>%
  filter(low_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, memory)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Cost Time Series",
       subtitle = "Memory Cost vs. Revenue from Low-Growth Companies") +
  ylab("Dollars per Megabyte") +
  xlab("Revenue (in billions)")
```

Computing Cost Time Series

Memory Cost vs. Revenue from Low-Growth Companies

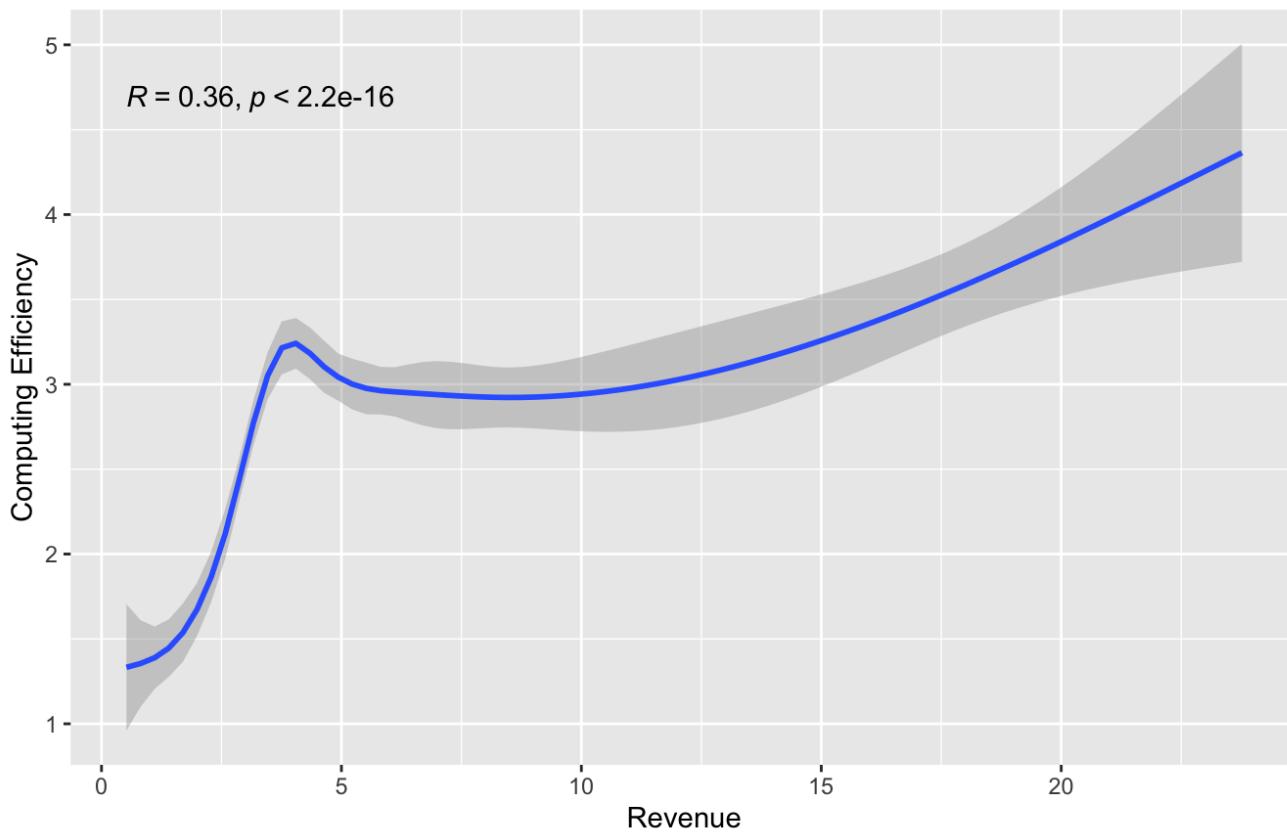


Unsuccessful Companies vs. Technological Improvements

```
# computing efficiency
fortune_joined_v4 %>%
  filter(unsuccessful == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, -log10(`Computing efficiency`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Efficiency Time Series",
       subtitle = "Computing Efficiency vs. Revenue from Less-Successful Companies") +
  ylab("Computing Efficiency")
```

Computing Efficiency Time Series

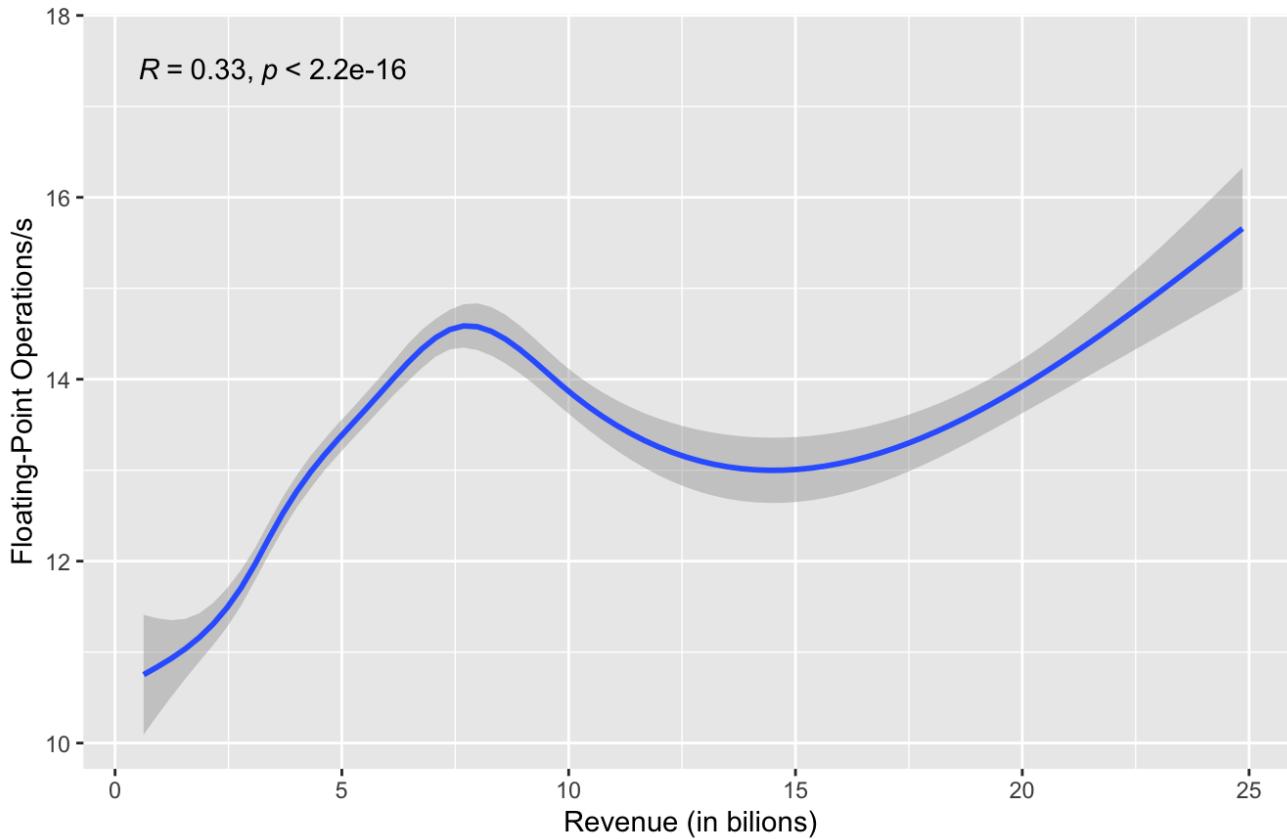
Computing Efficiency vs. Revenue from Less-Successful Companies



```
# supercomputer power
fortune_joined_v4 %>%
  filter(unsuccessful == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, log10(`Floating-Point Operations per Second`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Supercomputer Power Time Series",
       subtitle = "Floating-Point Operations/s vs. Revenue from Less-Successful Companies")
) +
  ylab("Floating-Point Operations/s") +
  xlab("Revenue (in billions)")
```

Supercomputer Power Time Series

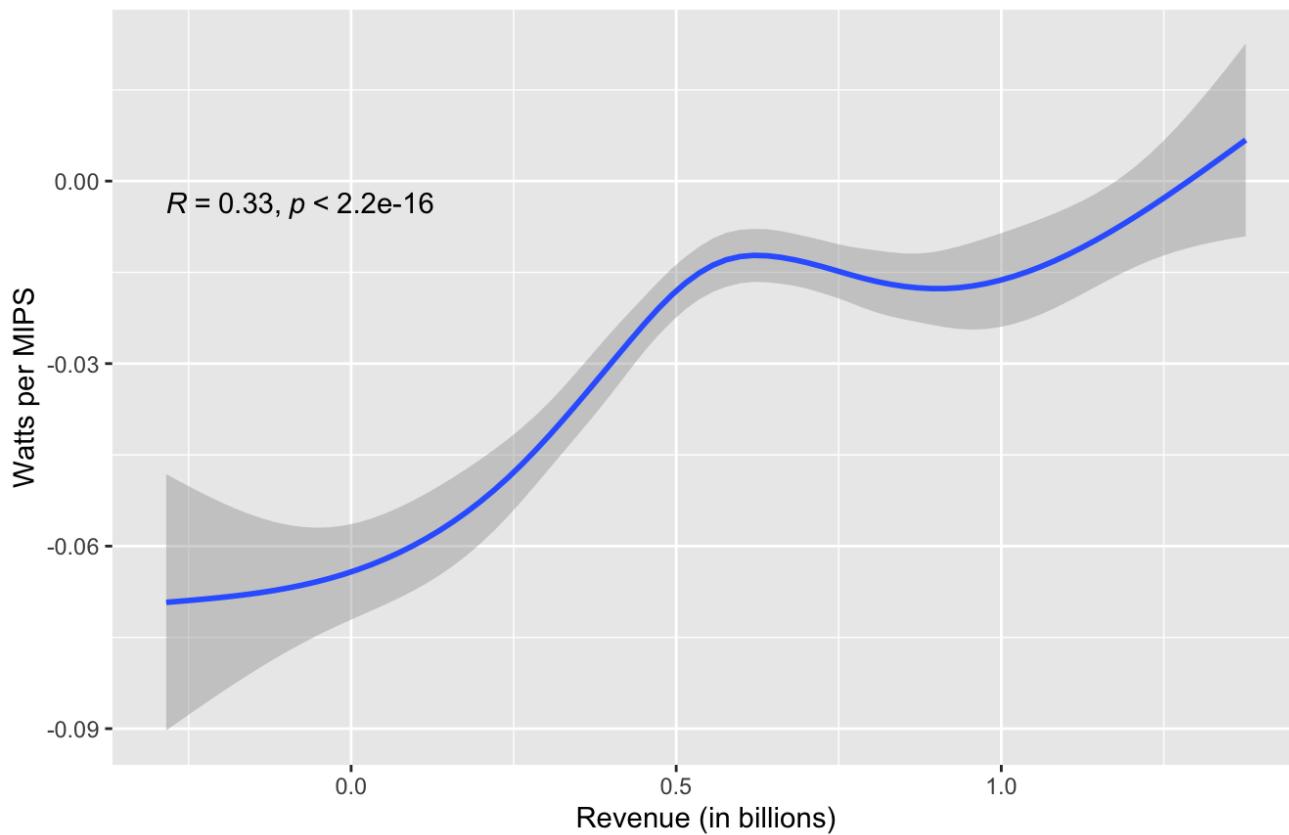
Floating-Point Operations/s vs. Revenue from Less-Successful Companies



```
# computing efficiency
fortune_joined_v4 %>%
  filter(unsuccessful == "True") %>%
  group_by(Year) %>%
  ggplot(aes(log10(Revenue), -`Computing efficiency`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Efficiency",
       subtitle = "Watts per MIPS vs. Revenue from Less-Successful Companies") +
  ylab("Watts per MIPS") +
  xlab("Revenue (in billions)")
```

Computing Efficiency

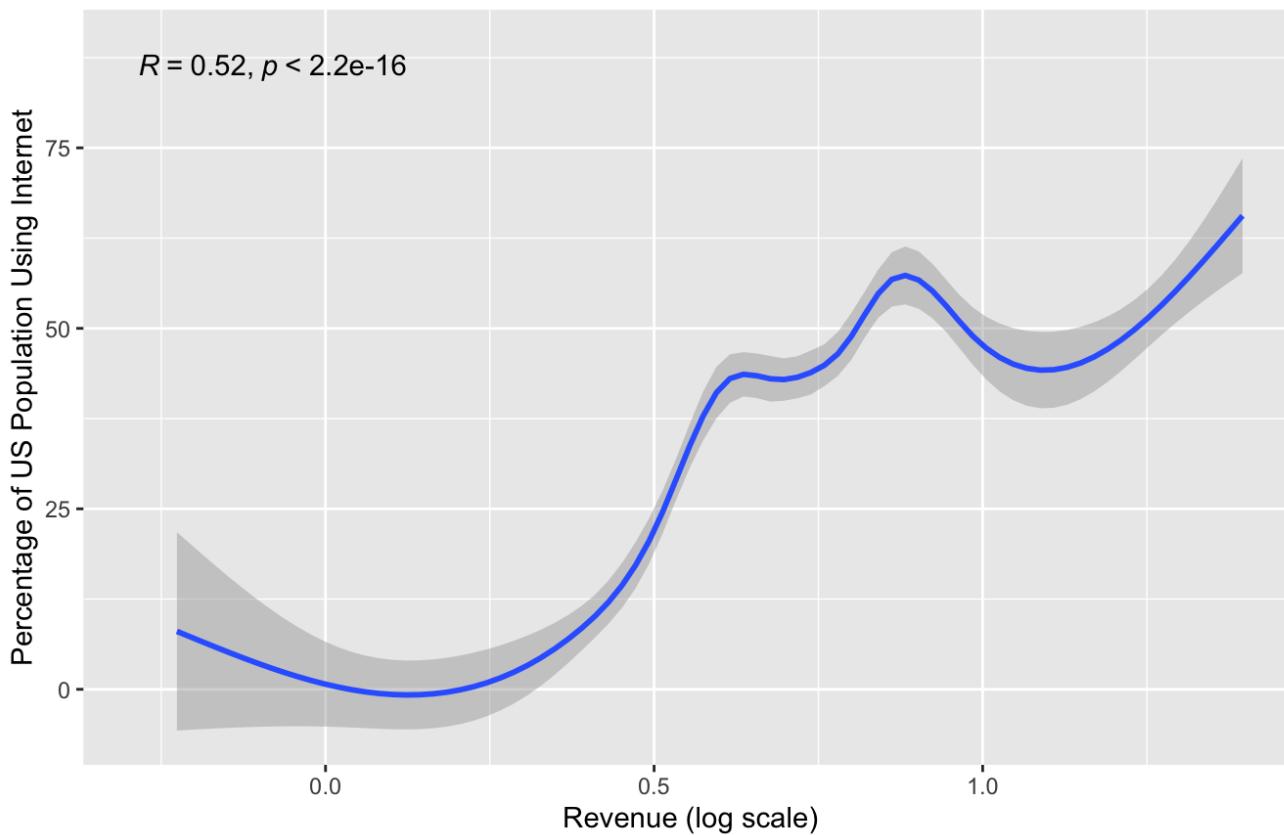
Watts per MIPS vs. Revenue from Less-Successful Companies



```
# percent of US using the internet
fortune_joined_v4 %>%
  filter(unsuccessful == "True") %>%
  group_by(Year) %>%
  ggplot(aes(log10(Revenue), `Individuals using the Internet (% of population)`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Internet Usage Time Series",
       subtitle = "Internet Usage vs. Revenue from Less-Successful Companies") +
  ylab("Percentage of US Population Using Internet") +
  xlab("Revenue (log scale)")
```

Internet Usage Time Series

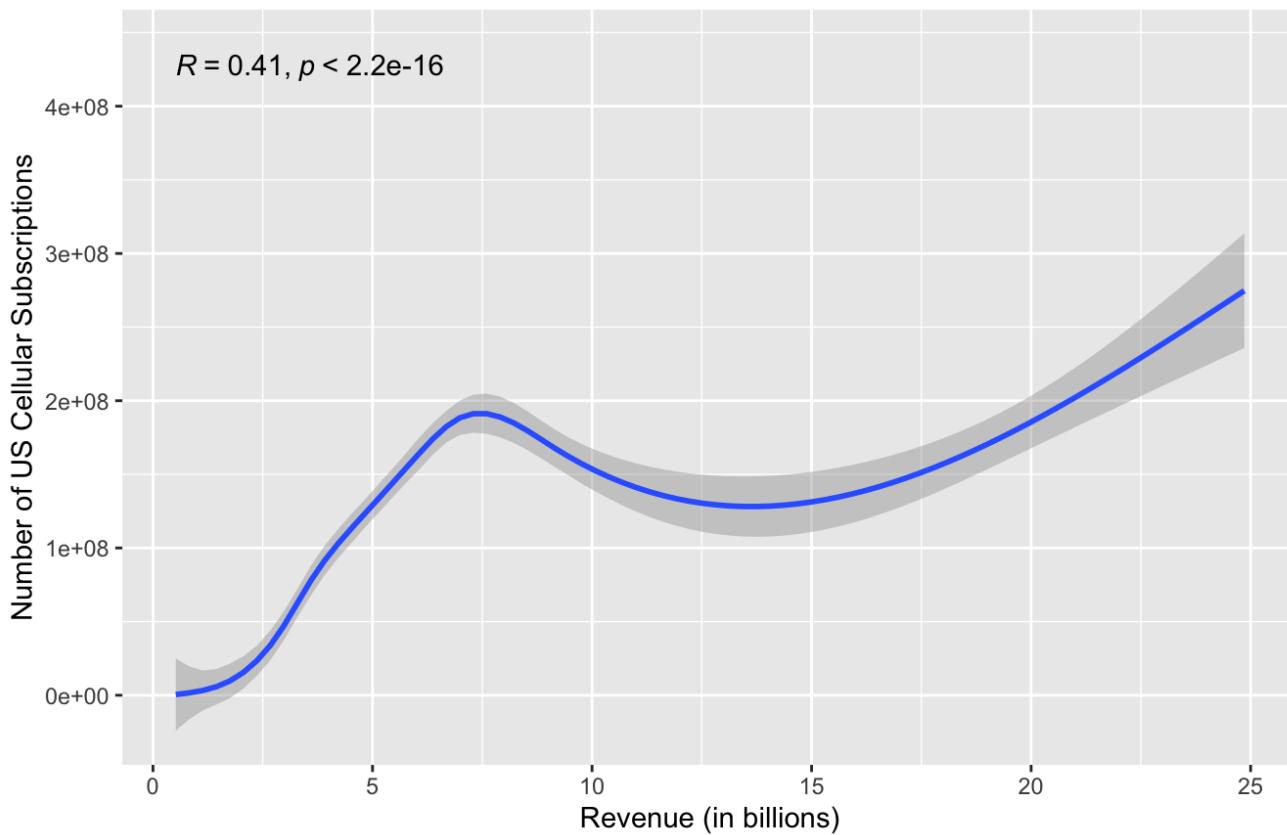
Internet Usage vs. Revenue from Less-Successful Companies



```
# Cellular Subscriptions
fortune_joined_v4 %>%
  filter(unsuccessful == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, `Cellular Subscriptions`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "US Cellular Subscriptions Time Series",
       subtitle = "Number of Cellular Subscriptions vs. Revenue from Less-Successful Companies") +
  ylab("Number of US Cellular Subscriptions") +
  xlab("Revenue (in billions)")
```

US Cellular Subscriptions Time Series

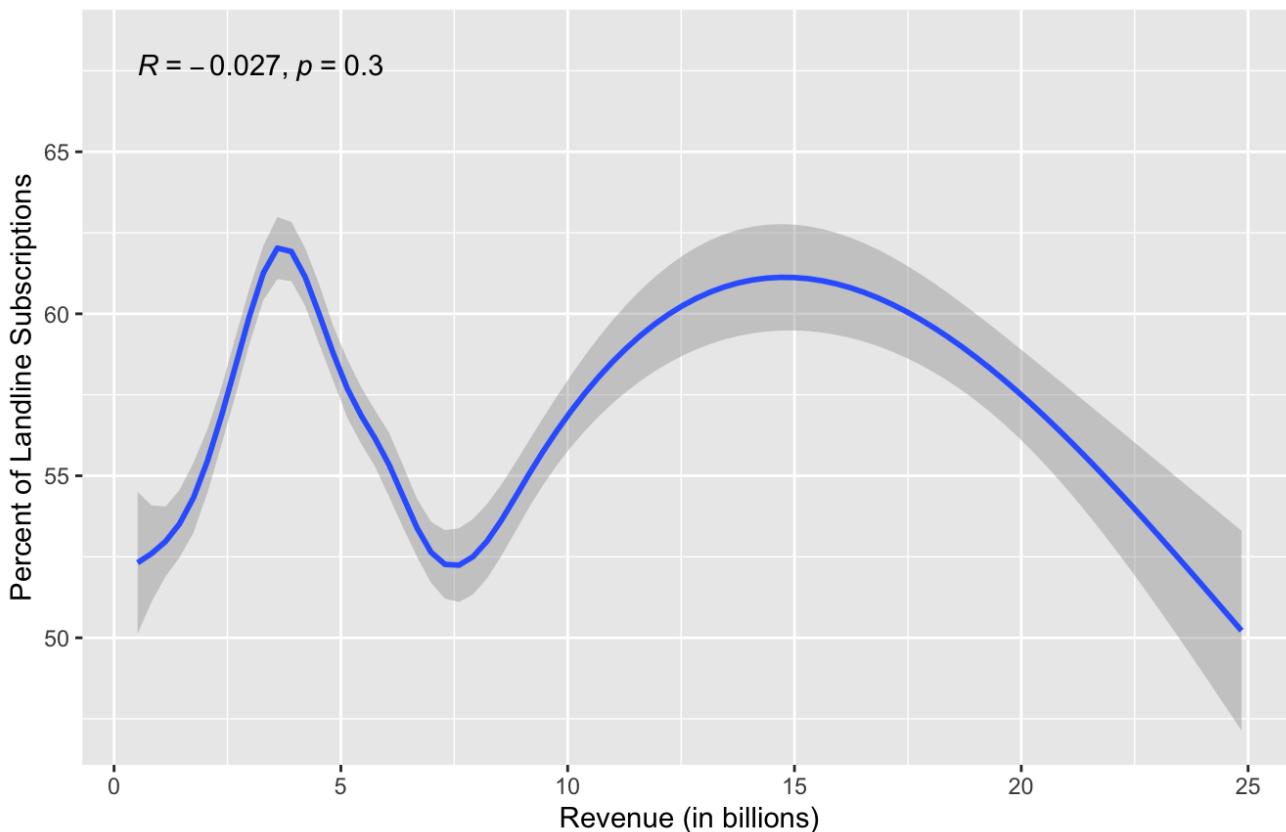
Number of Cellular Subscriptions vs. Revenue from Less-Successful Companies



```
# landline subscriptions (per 100 people)
fortune_joined_v4 %>%
  filter(unsuccessful == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, `Landline Subscriptions (per 100 people)`)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "US Landline Subscriptions Time Series",
       subtitle = "Percent of US using Landline vs. Revenue from Less-Successful Companies") +
  ylab("Percent of Landline Subscriptions") +
  xlab("Revenue (in billions)")
```

US Landline Subscriptions Time Series

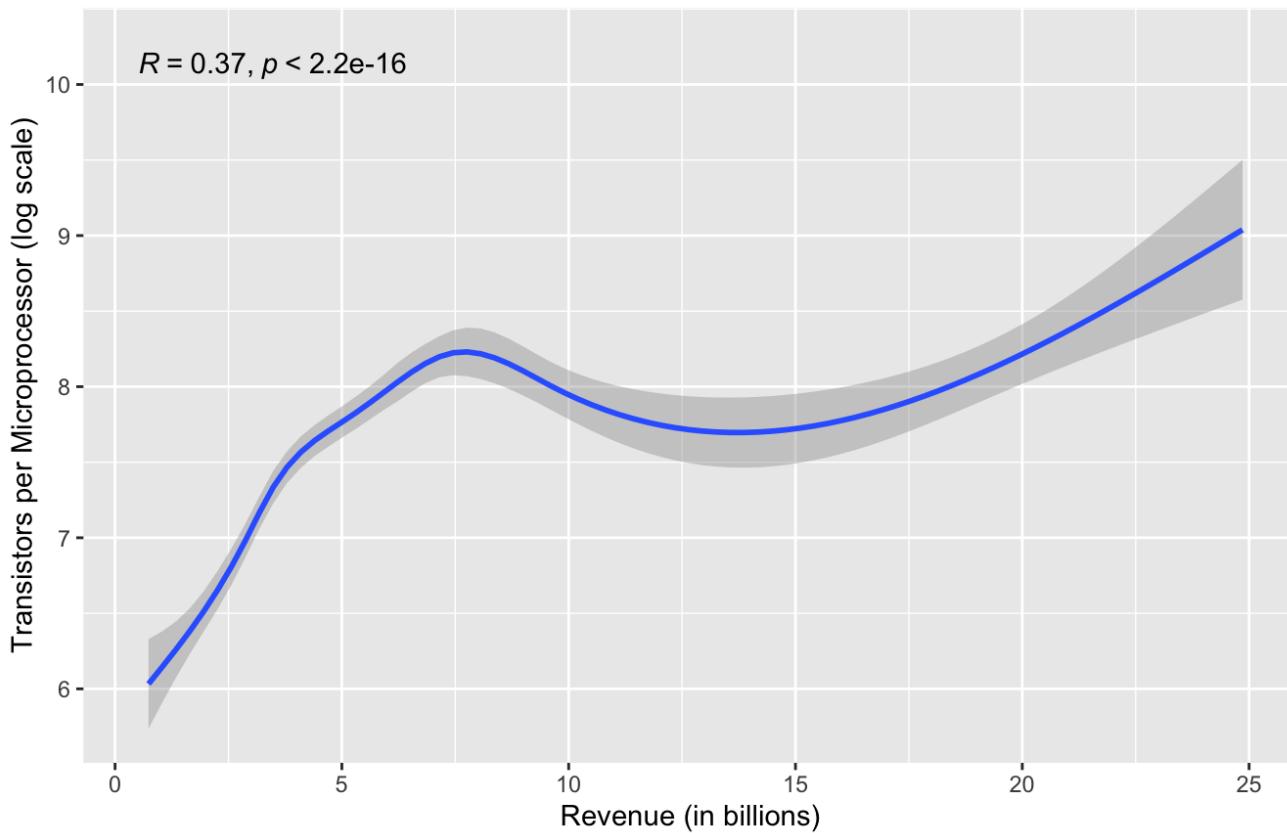
Percent of US using Landline vs. Revenue from Less-Successful Companies



```
# moores law
fortune_joined_v4 %>%
  filter(unsuccessful == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, log10(`Transistors per microprocessor`))) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Mooore's Law Time Series",
       subtitle = "Transistors per Microprocessor vs. Revenue from Less-Successful Companies") +
  ylab("Transistors per Microprocessor (log scale)") +
  xlab("Revenue (in billions)")
```

Mooore's Law Time Series

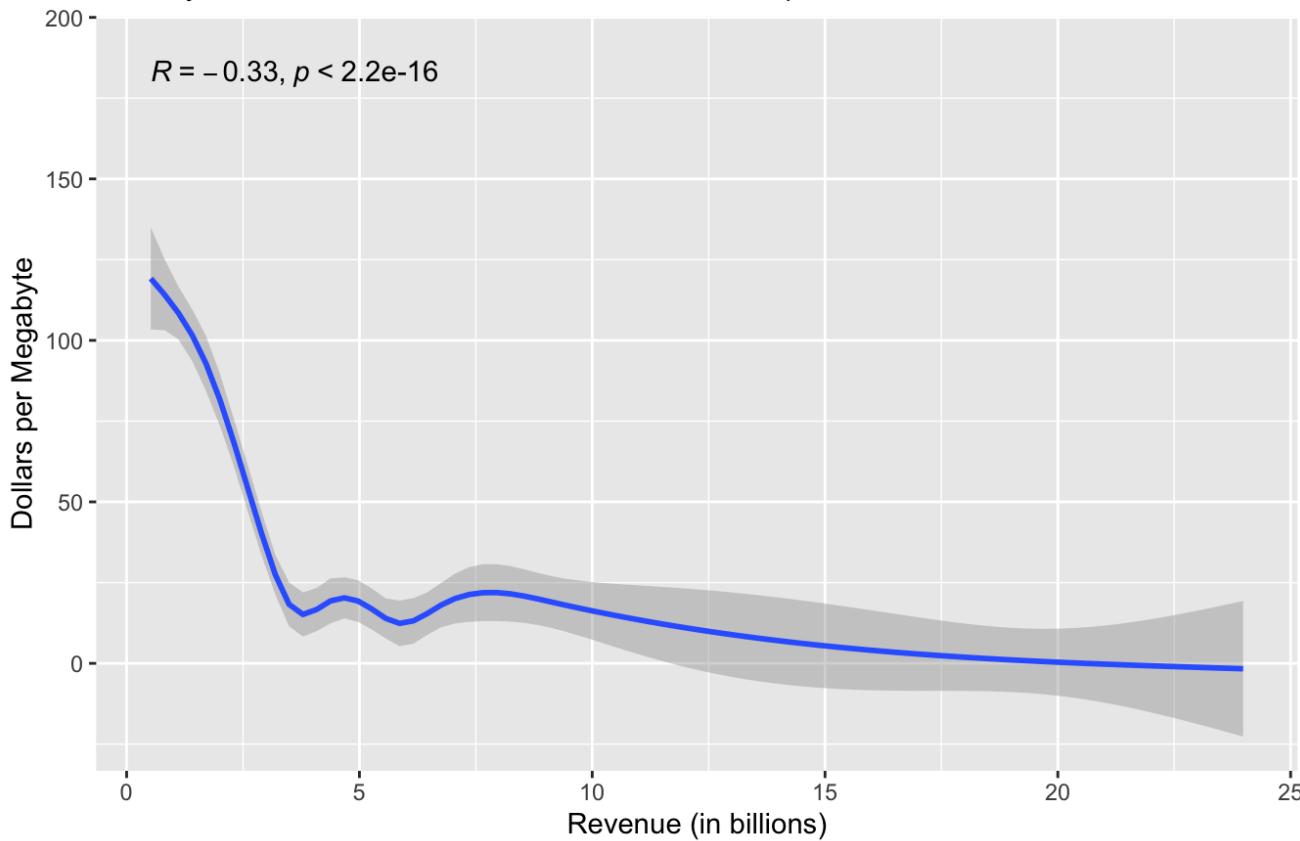
Transistors per Microprocessor vs. Revenue from Less-Successful Companies



```
# computing cost (memory)
fortune_joined_v4 %>%
  filter(unsuccessful == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, memory)) +
  geom_smooth() +
  stat_cor() +
  labs(title = "Computing Cost Time Series",
       subtitle = "Memory Cost vs. Revenue from Less-Successful Companies") +
  ylab("Dollars per Megabyte") +
  xlab("Revenue (in billions)")
```

Computing Cost Time Series

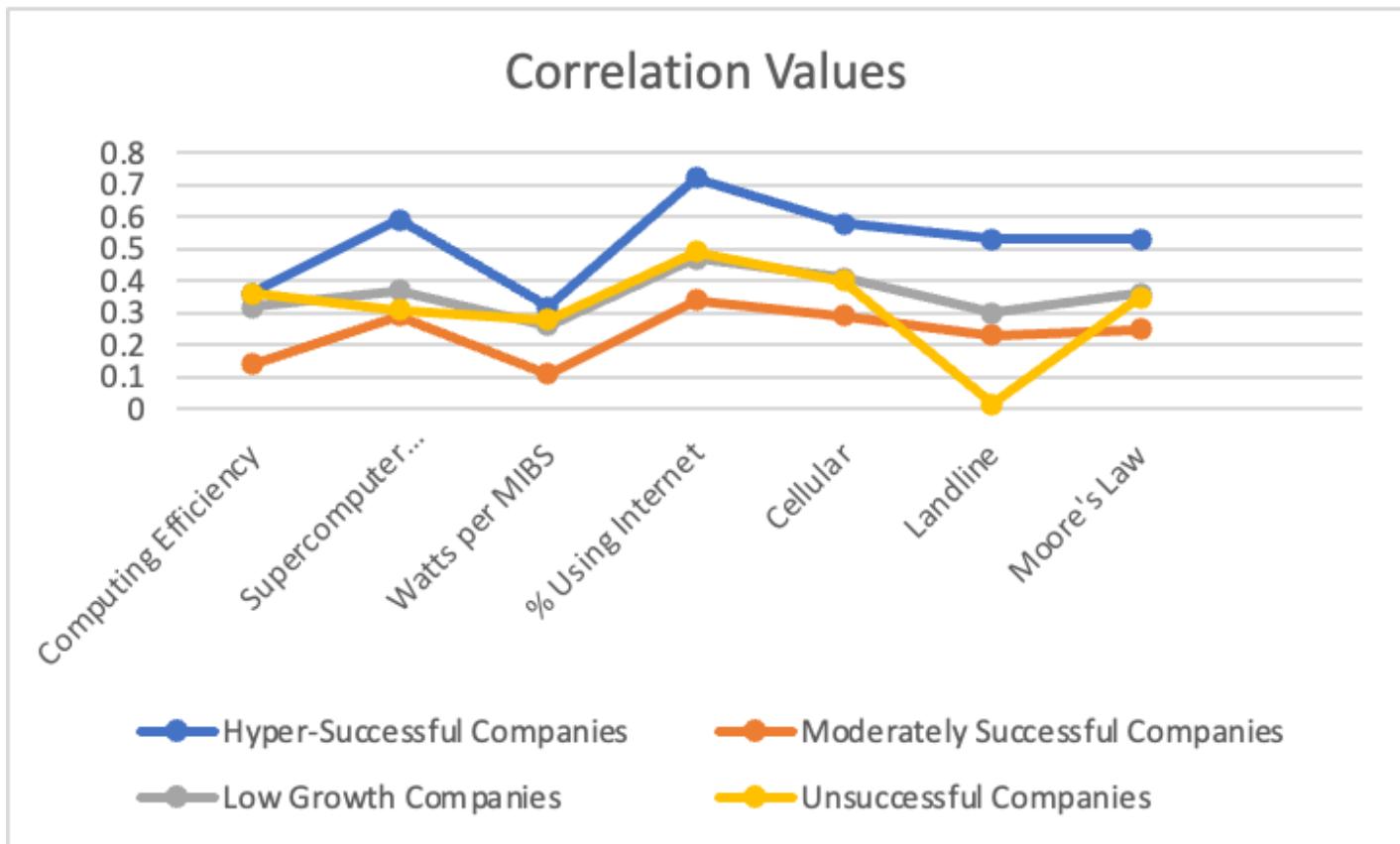
Memory Cost vs. Revenue from Less-Successful Companies



Correlation Results

Correlation Values

	Watts						Moore's Law	Computing Cost
	Computing Efficiency	Supercomputer Power	per MIBS	% Using Internet	Cellular	Landline		
Hyper-Successful Companies	0.36	0.59	0.32	0.72	0.58	-0.53	0.53	-0.18
Moderately Successful Companies	0.14	0.29	0.11	0.34	0.29	-0.23	0.25	-0.14
Low Growth Companies	0.32	0.37	0.26	0.47	0.41	-0.3	0.36	-0.25
Unsuccessful Companies	0.36	0.31	0.28	0.49	0.4	-0.016	0.35	-0.33



Interpretation:

- We can eliminate the following categories due to low overall correlation: *Computing Cost*, *Watts Per MIBS*, *Computing Efficiency*
- Supercomputer power shows significantly higher correlation to hyper-successful company growth compared to growth of even moderately successful companies. This suggest that Supercomputer power could be a technological advancement that highly successful companies have taken advantage of properly.
- Percent of US individuals Using the Internet also has much higher correlation to Hyper-Successful Companies.
- Cellular has a strong positive correlation with Hyper-Successful companies relative to the other categories.
- Landline is extremely negatively correlated, so disproportionately so that it may be worthwhile investigating why this could be the case.
- Moore's law shows promising correlation with Hyper-Successful companies as well.
- Low-Growth companies do not exhibit remarkably different performance from unsuccessful companies, so, for the sake of brevity, this subset may be ignored.

Correlation does not imply causation. However, the significance of these correlation values is inherent to the relative difference in correlation between hyper-successful companies and unsuccessful companies. There is certainly a trend in which correlated factors such as *Percent of US Population using Internet* display much stronger correlation to revenue of successful companies.

At first, I thought that the higher correlation was due to hyper successful companies having a stronger growth rate overall, resulting in those companies having a stronger correlation with technologies that increased in usage over time. This is not the case. The correlation was typically (in indicative cases) strong among hyper-successful companies, and weak among both moderately successful and unsuccessful companies. Considering how these

technologies do not experience directly increasing growth over the years, these correlation values suggest that hyper-successful companies take advantage of using new technologies (supercomputer power, percent using the internet, cellular, moore's law).

In short:

- Covariance analysis on the different company subsets suggests that new technology innovation and adaption impacts highly successful businesses more than other businesses.

In-Depth analysis of Hyper-Successful Businesses

The previous analysis suggest that technological advancements and adoption disproportionately affected hyper-successful companies relative to moderately successful companies. This is not unexpected, but it is interesting. Additionally, it also is interesting how the technologies related in improving hardware were much less correlated with success, but the technology advancements related to technology distribution (phone + internet usage) and physical implementation (supercomputer power) had a higher impact on successful companies.

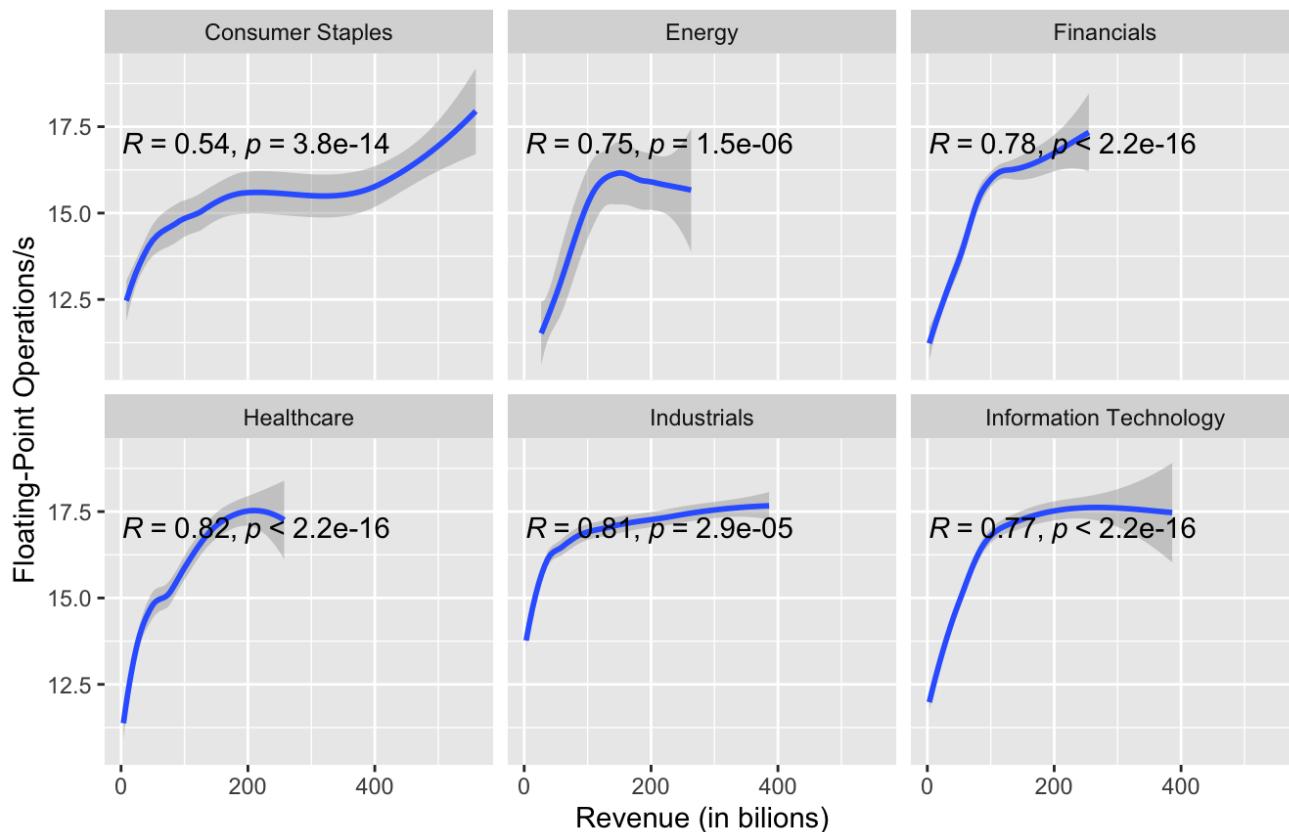
This suggests that these high-performance companies have a good sense of how consumer trends behave, and have taken the right moves to take advantage of different consumer trends, such as increasing online advertising during the period of rapid growth of the new internet.

The final goal of this analysis is to determine what role sector plays within these highly successful businesses. We will be ignoring *Computing Cost*, *Watts Per MIBS*, *Computing Efficiency* as they did not display conclusive correlation one way or another among the greater population. Additionally, landline subscriptions does not show promising direction, so I will not plot that either (I also am trying to eliminate excess plots, as I already know this will be a pain to grade, and I'm sorry about that).

```
# supercomputer power
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, log10(`Floating-Point Operations per Second`))) +
  geom_smooth() +
  stat_cor() +
  facet_wrap(~ Sector) +
  labs(title = "Supercomputer Power Time Series",
       subtitle = "Floating-Point Operations/s vs. Revenue from Hyper-Successful Companies") +
  ylab("Floating-Point Operations/s") +
  xlab("Revenue (in billions)")
```

Supercomputer Power Time Series

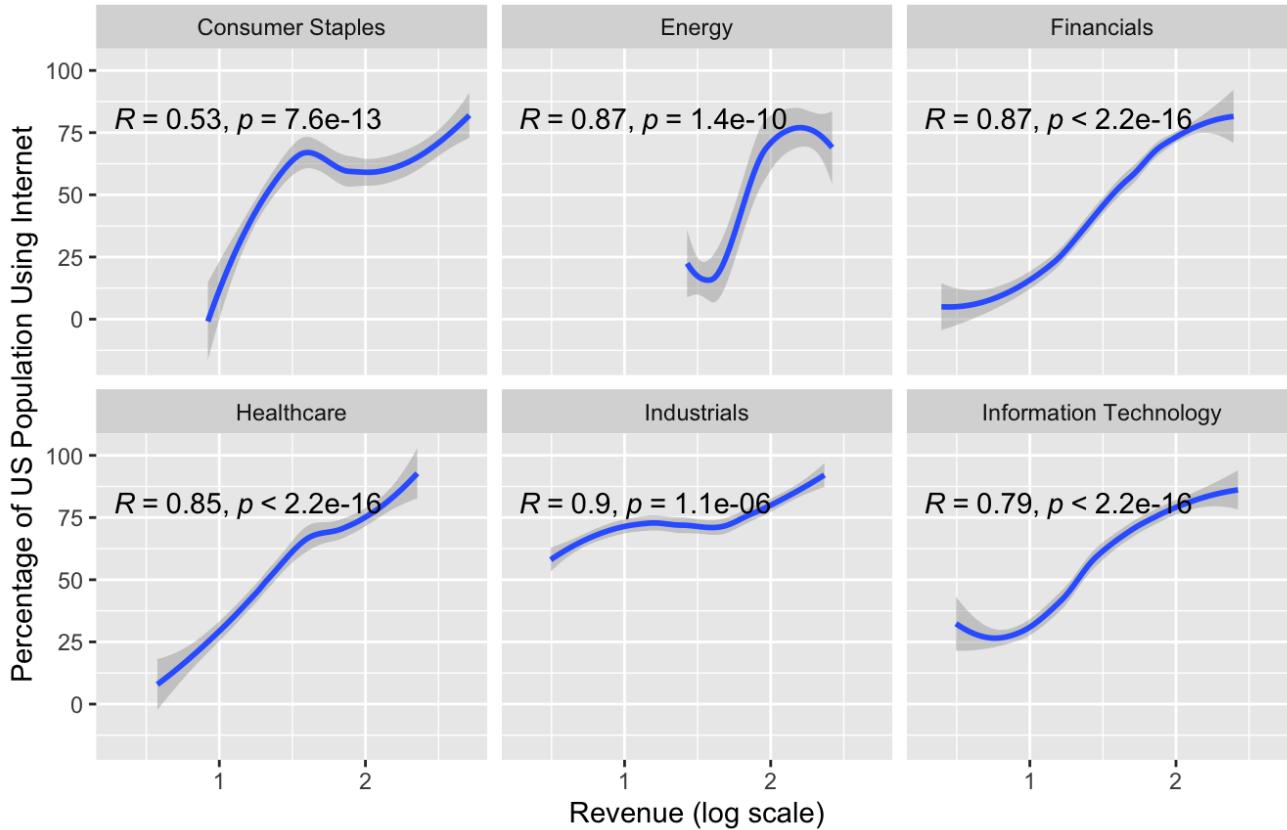
Floating-Point Operations/s vs. Revenue from Hyper-Successful Companies



```
# percent of US using the internet
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(log10(Revenue), `Individuals using the Internet (% of population)`)) +
  geom_smooth() +
  stat_cor() +
  facet_wrap(~ Sector) +
  labs(title = "Internet Usage Time Series",
       subtitle = "Internet Usage vs. Revenue from Hyper-Successful Companies") +
  ylab("Percentage of US Population Using Internet") +
  xlab("Revenue (log scale)")
```

Internet Usage Time Series

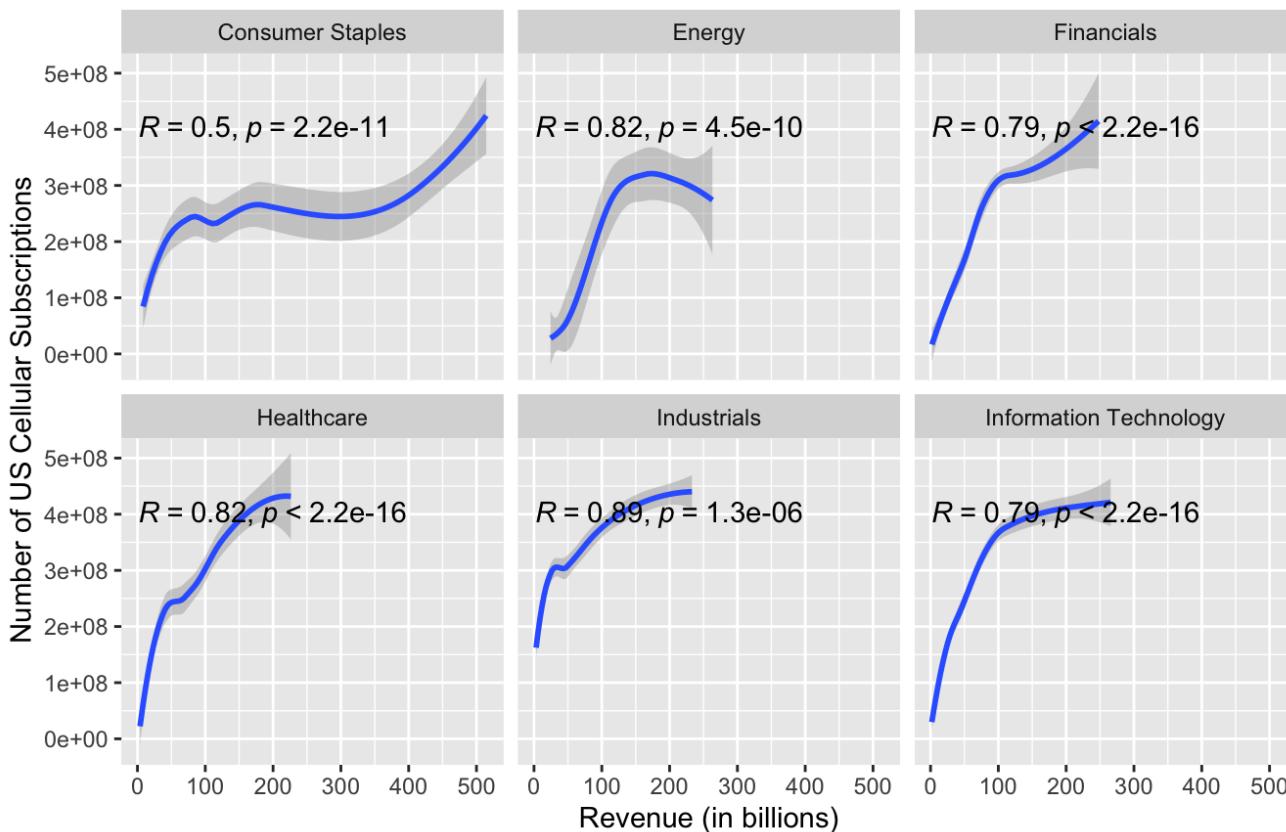
Internet Usage vs. Revenue from Hyper-Successful Companies



```
# Cellular Subscriptions
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, `Cellular Subscriptions`)) +
  geom_smooth() +
  stat_cor() +
  facet_wrap(~ Sector) +
  labs(title = "US Cellular Subscriptions Time Series",
       subtitle = "Number of Cellular Subscriptions vs. Revenue from Hyper-Successful Companies") +
  ylab("Number of US Cellular Subscriptions") +
  xlab("Revenue (in billions)")
```

US Cellular Subscriptions Time Series

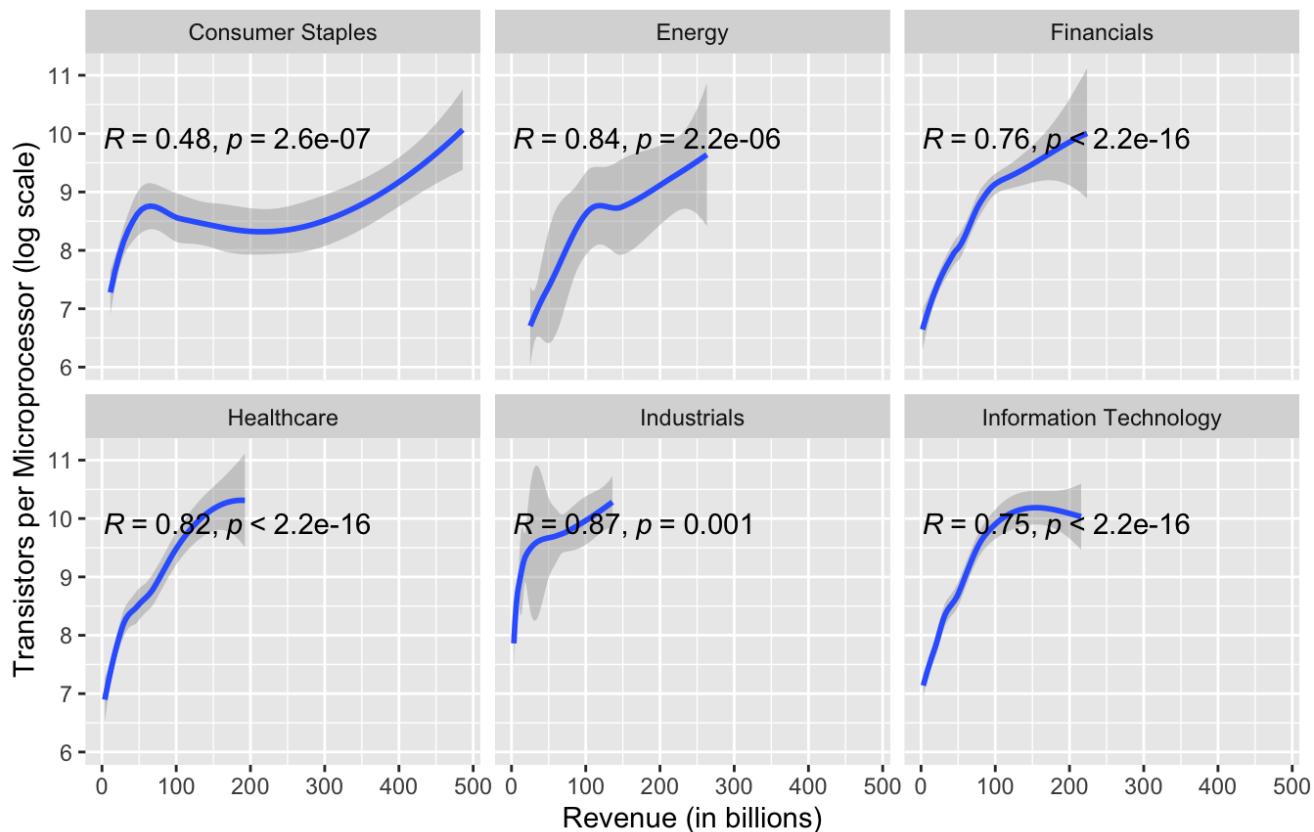
Number of Cellular Subscriptions vs. Revenue from Hyper-Successful Companies



```
# moores law
fortune_joined_v4 %>%
  filter(hyper_growth == "True") %>%
  group_by(Year) %>%
  ggplot(aes(Revenue, log10(`Transistors per microprocessor`))) +
  geom_smooth() +
  stat_cor() +
  facet_wrap(~ Sector) +
  labs(title = "Mooore's Law Time Series",
       subtitle = "Transistors per Microprocessor vs. Revenue from Hyper-Successful Companies") +
  ylab("Transistors per Microprocessor (log scale)") +
  xlab("Revenue (in billions)")
```

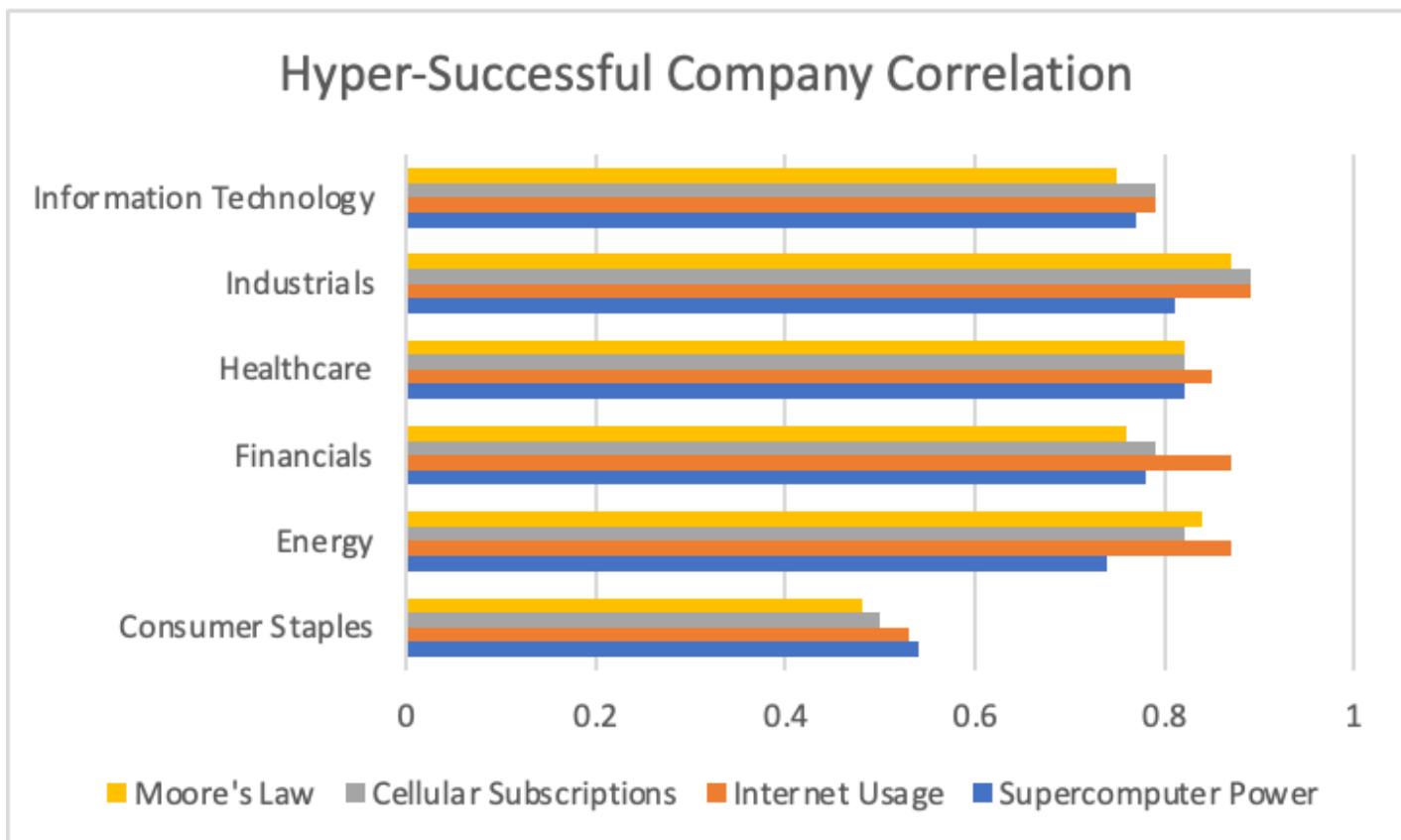
Mooore's Law Time Series

Transistors per Microprocessor vs. Revenue from Hyper-Successful Companies



Correlation for Successful Companies, by Sector

	Consumer Staples	Energy	Financials	Healthcare	Industrials	Information Technology
Supercomputer Power	0.54	0.74	0.78	0.82	0.81	0.77
Internet Usage	0.53	0.87	0.87	0.85	0.89	0.79
Cellular Subscriptions	0.5	0.82	0.79	0.82	0.89	0.79
Moore's Law	0.48	0.84	0.76	0.82	0.87	0.75



Breaking down the analysis to the sector scale provides a better look into which sectors benefit more from various technological changes. It is relatively evident that there is a correlation between more high tech sectors (i.e., not Consumer Staples) and success due to technological advancements. This is an expected result since consumer staples companies rely much less heavily on innovation, relative to companies in other sectors.

Modeling Hyper-Successful Companies

I will now create a model of Hyper-Successful companies that can be used to predict (to some degree) where we could expect to see hyper-successful ending up in the next few years.

```
# log transform revenue to create linear pattern

# model of Revenue per year for hyper successful companies

# plotting revenue of fortune 500 as a whole with hyper-successful

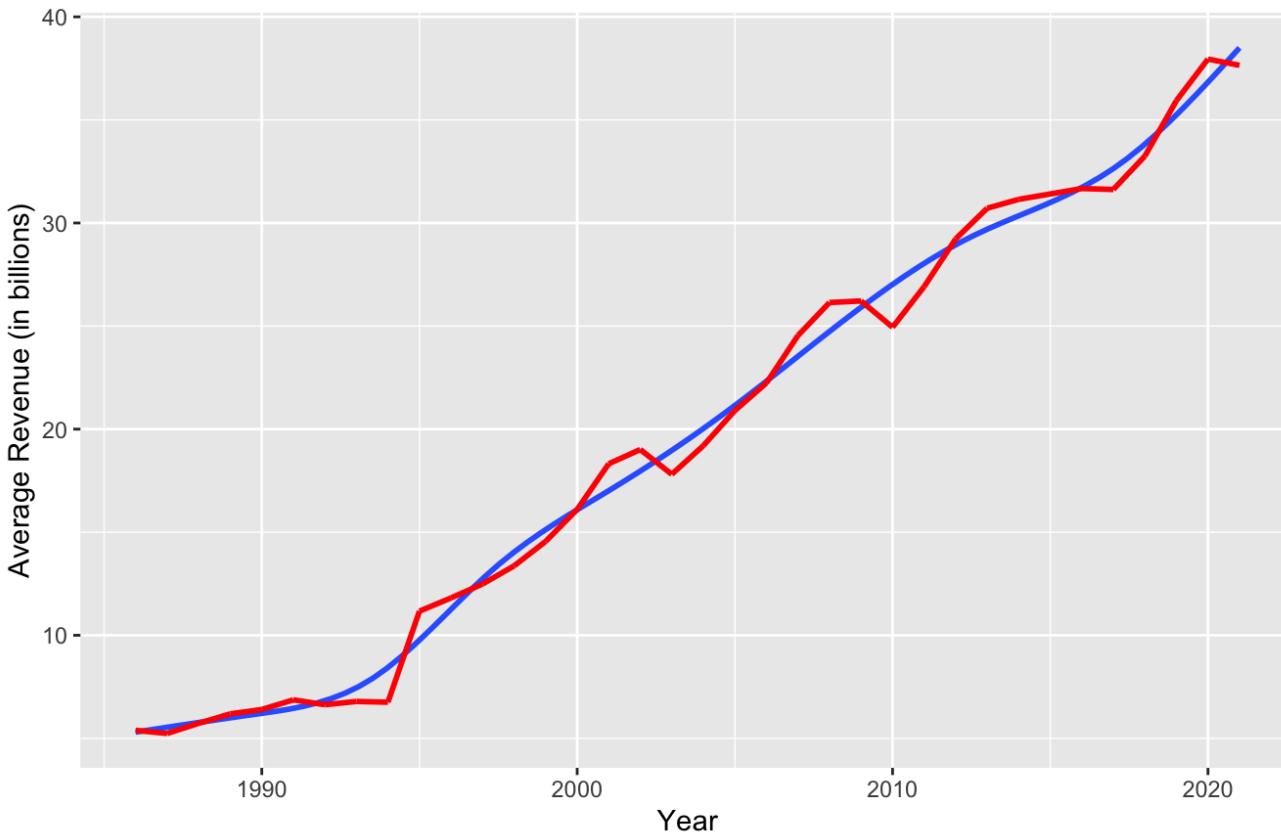
fortune_joined_v5 <- fortune_joined_v4 %>%
  group_by(Year) %>%
  mutate(meanrev = mean(Revenue, na.rm = TRUE))

hyper_model <- lm(Year ~ meanrev, data = fortune_joined_v5)

plot <- fortune_joined_v5 %>%
  add_predictions(hyper_model, "Revenue_pred")

ggplot(fortune_joined_v5, aes(Year, meanrev)) +
  geom_smooth() +
  geom_line(data = plot, color = "Red", size = 1) +
  labs(title = "Model of Hyper Successful Revenue",
       subtitle = "Year vs. Revenue from Hyper-Successful Companies") +
  ylab("Average Revenue (in billions)")
```

Model of Hyper Successful Revenue
Year vs. Revenue from Hyper-Successful Companies



```
variable_years<-data.frame(meanrev=c(2022,2023,2024,2025,2030,2040))

# produce predictions for 2022-2040
predict(hyper_model, newdata = variable_years)

##          1         2         3         4         5         6
## 3950.545 3951.517 3952.490 3953.462 3958.324 3968.048
```

Summary of model, which can be used to make future predictions about revenue of hyper successful companies.

```
summary(hyper_model)
```

```
##
## Call:
## lm(formula = Year ~ meanrev, data = fortune_joined_v5)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.5666 -0.8210  0.0482  0.8774  3.1078
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.984e+03 1.892e-02 104874    <2e-16 ***
## meanrev     9.724e-01 8.558e-04    1136    <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.269 on 20918 degrees of freedom
## Multiple R-squared:  0.9841, Adjusted R-squared:  0.9841
## F-statistic: 1.291e+06 on 1 and 20918 DF,  p-value: < 2.2e-16
```

Part 6: Summary & Results

How did the rise in technology usage affect historically successful businesses/industries?

This analysis has appropriately examined and evaluated how the rise in technology usage has impacted historically successful businesses & industries.

The first task of preparing the Fortune 500 dataset was tedious, as it was difficult to find data to identify sectors of every company. I managed to find a dataset holding 7 million company records. After wrangling that data together along with the other datasets, I was able to start the analysis. It quickly became evident that I needed to recode the dataset somehow to reduce the categories down to a tangible count.

I had already planned on using stock market sectors as a way of grouping companies. Sectors make sense for many reasons, one of the largest being that their stock prices often trade in similar direction, so numerical values for these companies should maintain a certain level of indifference from the confounding variable of market price. This indifference is due to the fact that when grouping by sector, that entire sector will be effected almost entirely the same (in aggregate) by external market stimuli. For the purpose of analyzing the effects of technology, it is important to reduce any external stimuli that could introduce confounding factors to our revenue values.

This analysis suggest that hyper-successful companies can attribute some of their success to technology. The rate of distribution/improvements in certain technologies (such as internet usage) display much higher correlation with hyper-successful companies revenue compared to other companies. Revenue trends of hyper-successful businesses suggest that with the disruption of new technology, only the top companies truly manage to reap the entire benefits of such disruptive change.

Limitations

Firstly, it would have been nice to be able to find Fortune 500 data for 2015. I don't know why that data does not exist, but for the purposes of my analysis, it isn't absolutely necessary. Additionally, further examination of potential confounding variables would clarify these results, but I didn't want to go overboard on the analysis.