COMP3042 Coursework

Name: Goh Xin Yee

Student ID: 20093715

Date of Completion: 14/4/2022

Goal & Objectives: This project aims to visualize key data about **Apple,** giving insights into Apple's business performance over the years.

Problem Statement: As one of the tech giants, how's Apple performing over the years?

Target Audience: People who are interested in technology, business, economics, or even Apple fans.

Description of data (D)

D1: Apple Revenue by Quarter
Records the quarterly revenues of Apple from 2006 to 2021.

D2: Apple Revenue by Region
Records the revenue generated by 5 different regions – the Americas, Europe, China, Japan, and the Rest of Asia Pacific – for Apple from 2015 to 2021.

D3: Apple Revenue Sales by Product
Records the revenues and sales of Apple products, which are divided into 4 categories – iPhone, iPad, Mac, and Wearables & Home.

D4: Apple iPhone Price Evolution
Records the Start and End price of each iPhone model over the years – in USD and MYR currencies.

D5: Apple Stores Worldwide
Records the locations – by regions – of Apple stores around the world with the respective number of stores.

D6: Apple Global Market Share 2021

Records the share(%) of different smartphone brands in the Global Smartphone Market 2021. There are a total of 9 brands – Samsung, Apple, Huawei, Xiaomi, Motorola, LG, Vivo, Realme, OPPO – and two other categories known as “Others” and “Unknown”.

Description of questions (Q)

4 initial questions:

Q1: How’s the overall performance of Apple over the years?
Q2: How’s the performance of Apple across different regions in the world?
Q3: How’s the distribution of Apple stores worldwide?
Q4: How’s the performance of Apple products over the years?

6 further questions:

Q5: Which region has contributed the most to Apple’s business?
Q6: Which Apple product has generated the most revenue for Apple over the years?
Q7: What is the evolution of iPhones’ prices over the years?
Q8: Which iPhone is the most expensive model and which are the more budget-friendly model(s)?
Q9: How are the sales of iPhones with the changing price over the years?
Q10: How was the performance of Apple in 2021 from a global perspective?

Description of visualization strategies

Link to Shiny app: https://n9ss3a-xinyee-goh.shinyapps.io/AppleInfoVisualization/
This section will discuss the data representations used to **answer each question** and the corresponding procedures. *ggplot2*, *plotly*, and *shiny* are used for plotting and interactivity. The outputs of each plot can be accessed in **Appendix, click the question hyperlink to go to the respective plots**.

Q1: How's the revenue of Apple over the years?

- Plot **a line graph** to showcase Apple's revenue by year.
- Plot **a stacked/grouped bar chart** to showcase Apple's revenue by quarter.
- Both from 2006 to 2021, using **D1**.
- User can switch to view different graphs.

Procedures:

1. As dataset D1 records only the quarterly revenues, group the dataset by Year and find the Total Revenue for each year.
2. For the line graph, encode Year and Total Revenue using **position** on the x-axis and y-axis, respectively.
3. For stacked/grouped bar chart, encode Year using **position** on the x-axis, Quarterly Revenue using **size** (length of the bar), and Quarter using **colours**.

Q2: How's the performance of Apple across different regions in the world?

- Plot **a stacked/grouped bar chart** to showcase Apple's revenue by region, from 2015 to 2021, using dataset **D2**.

Procedures:

1. Encode Year using **position** on the x-axis, Revenue using **size** (length of the bar), and Region using **colours**.

Q3: How's the distribution of Apple stores worldwide?

- Plot a **map** to showcase Apple's store locations around the world and the corresponding number of stores, using **D5**.

Procedures:

1. Get the *world map* from R, which records the longitude and latitude of regions across the world.
2. Dataset D5 contains region names and the corresponding Number of Stores. Left join the *world map* with D5 by common Region. The joined dataset will have Region, Longitude, Latitude, and Number of Stores.
3. To plot the map using the joined dataset, encode Latitude and Longitude using **position** on the x and y-axis respectively, and the Number of Stores using **colours**.

Q4: How's the performance of Apple products over the years?

- Plot **a bar-line chart with a dual y-axis** to showcase the revenue and sales of different Apple product categories over the years, using **D3**. One y-axis for Revenue, one y-axis for Sales. The year range is based on the year the product category was first launched, till 2021.
- **Each product category will have 1 bar-line chart** – a small-multiple strategy. Users can choose the product category to view.

Procedures:

1. Group the dataset by Product Category.
2. Filter out the rows of data where the Product Category is the selected category.
3. Choose distinct data representations for Revenue and Sales – bar for revenue, line for sales.
4. To plot, encode Year using **position** on the x-axis, Revenue using **size** (length of bars) and Sales using **position** (of line) on the y-axis.
5. Revenue and Sales are also encoded using **colour** for distinction.
6. Do the same for other product categories when selected by users.

Q5: Which region has contributed the most to Apple's business?

- Plot **a pie chart** to showcase the revenue contribution (%) of each region from 2015 to 2021, using **D2.**
- Each year will have 1 pie chart – a small-multiple strategy. Users can choose the year to view.

Procedures:

1. Filter out the rows of data where the Year is the selected year.
2. Calculate the Revenue Contribution (%) of each region for the selected year using the formula:

   `Revenue of a Region/sum(Revenue)*100.`
3. Mutate the dataset by adding a new variable to record the Revenue Contribution(%). The mutated dataset will now contain 4 fields: Year, Region, Revenue, and Revenue Contribution(%).
4. To plot, encode Region using **colours** and Revenue Contribution(%) using **size**.
5. Do the same for other years when selected by users.

Q6: Which Apple product has generated the most revenue for Apple over the years?

- Plot **an area graph** to showcase the revenue contribution(%) of each product category to Apple's revenue from 2015 to 2021, using **D3**. The year range is selected for the reason that 2015 is when each category has had at least 1 product launched.

Procedures:

1. Factorise Product Category – iPhone, iPad, Mac, Wearables&Home – as levels from top to bottom.
2. The current dataset contains data on products that were launched **before** 2015, which don't include all 4 categories. Hence, subset the dataset to get a new dataset that contains only the data **starting in 2015 and beyond**.
3. Group the new dataset by Year.
4. Calculate the Total Revenue generated by all categories, for each year.
5. Calculate the Revenue Contribution(%) of each category in each year based on the corresponding Total Revenue. Mutate the dataset to record the Revenue Contribution(%).
6. To plot, encode Year using **position** on the x-axis, Revenue Contribution(%) using **size** (area coverage), and Product Category using **colours**.

Q7: What is the evolution of iPhones' prices over the years?

Q8: Which iPhone is the most expensive model and which is the more budget-friendly model(s)?

- Plot **a point range graph** to showcase the price evolution of iPhones from the 1st to the latest model, using **D4**.

Procedures:

1. Arrange the dataset by the year of release of the iPhone in ascending order.
2. Factorise the year of release into discrete levels, from the earliest year (top-level) to the latest year (bottom-level).
3. Factorise iPhone Model into levels, from the earliest model (top-level) to the latest model (bottom level).
4. Group the dataset by iPhone Model.
5. Calculate the Middle Price – median – of each model based on the provided start and end prices. Mutate the dataset to record the Middle Price.
6. To plot, encode iPhone Model using **position** on the x-axis, Start, End, and Middle Price using **colours** and **position** on the y-axis. The price range is encoded using **size** (length).

Q9: How are the sales of iPhones with the changing prices over the years?

- Plot **a scatter plot t**o showcase the relationship between iPhone sales and price using **D3** and **D4.** The average price of iPhones released in the same year is used to plot against the sales of iPhones in the corresponding year.

Procedures:

1. Filter dataset D3 to get the data of **iPhone** only.
2. Select only Year and Sales from the filtered dataset D3.
3. Group dataset D4 by iPhone model.
4. Calculate the Middle Price of each model based on the provided start and end prices. Mutate the dataset to record the Middle Price.
5. Group the mutated dataset D4 by Year.
6. Calculate the Average Price of the iPhone for each year based on the Middle Price of the iPhone models **released in the same year.** Mutate the dataset to record the Average Price.
7. Select only Year and Average Price from the mutated dataset D4.

8. Left join the modified dataset D3 and D4 by common Year. The joined dataset contains 3 fields: Year, Sales, and Average Price.
9. Since there were times when Apple released multiple iPhones in a year, the joined dataset will have rows with the same year and data. So, perform data cleaning on the dataset by selecting only the rows of data with distinct Year; drop the rest.
10. To plot, encode Average Price and Sales using **position** on the x and y-axis, respectively.
11. The best fit line is plotted to show the trend.

Q10: How was the performance of Apple in 2021 from a global perspective?

- Plot **a multi-line graph** to showcase the share(%) of Apple and other brands in the Global Smartphone Market 2021, using **D6**.
- Plot **a pie chart** to showcase the share(%) in month-based specific – a small-multiple strategy. Users can choose the month to view.

Procedures for multi-line graph:

1. Factorise each Date in the dataset into discrete Month-based levels.
2. To plot, encode Month and Share(%) using **position** on the x and y-axis, and Brands using **colours**.

Procedures for pie chart:

1. Subset the dataset to get a new dataset where the Month is the selected month.
2. To plot, encode Brands using **colours** and Share(%) using **size**.
3. Do the same for other months when selected by users.

The Exploratory Process and Findings

Generating Questions:

When planning the initial questions, I started by examining the datasets and considering the information that can be extracted from them to generate the initial questions. This process was similar to reverse engineering.

After visualizing data for the initial questions, those questions were revisited to make sure that they are fulfilled by the plots. Then from there, I further brainstormed ideas by asking: *What else can I know from these plots? What else do I want to know about Apple?* and derived more questions. As can be seen, the further questions are relatively related to the initial questions.

Generating Visualizations:

Before planning visualizations, **the message or goal of each question** was made clear to determine the data needed for the question. **The fitness of the datasets** was then accessed. This process has also helped in planning data cleaning and transformation.

Then, **data dimensions** were determined to know whether the target plots involve univariate or multivariate data. This has helped to decide the types of plots to be used for representing data and also prevented the involvement of too many variables in a single plot, which might make the plot overcomplicated to understand.

After that, the **data types** of the involved variables were accessed to check whether they are nominal, ordinal, or quantitative. Also, **their importance** was accessed to know what to focus on when plotting. This has helped to decide the visual encodings to use on the variables and their arrangement on the plots – e.g. whether the plots can be subdivided by certain variables and how to do so.

Finally, I explored ways to **refine the plots** by adding descriptive text, defining the scale and colour scheme, adding interactivity, etc.

Findings (Answers for Questions):

A-Q1: Apple's revenue experienced a significant increase starting in the year 2010. Its best performance was last year – 2021 – when it had obtained revenue of *$365.6billion USD*. Based on the quarterly graph, Apple's performance in terms of revenue is often better during the first quarter.

A-Q2: Overall, Apple's revenues have increased in each region over the years but there was a significant drop in revenue in the Asia Pacific in 2018 from *$44.1billion* to *$17.4billion.*

A-Q3: Apple currently has 517 stores worldwide. Apple hasn't expanded its business to Africa. The Americas is home to over 200 Apple stores, which makes it the Top 1 region with the most Apple stores.

A-Q4: iPhone experienced significant growth from the year 2007 to 2015, then underwent a drop from the year 2015 to 2017, and from 2018 to 2020. For **iPad**, there was an increase in sales from the year 2011 to 2013. However, its performance underwent a significant *trough* between the years 2013 and 2021. The sales of **Mac** have been steady, despite a slight drop in the years 2013, 2016, and 2019. **Wearables & Home** have experienced significant growth in sales from the year 2017 to 2021.

A-Q5: Americas has contributed the most to Apple's revenue from 2015 to 2021, with 40% and above. In 2021, its revenue contribution was 41.9%, which is equivalent to *$153.3billion USD*.

A-Q6: iPhone has been the main revenue generator for Apple over the years, with more than 60% of the contribution. iPhone's best contribution was in the year 2018 when it was responsible for 73.1% of the revenue.

A-Q7: As observed, the price of iPhone has been rising in general, except for a few models such as 'iPhone 5c', 'iPhone SE 1st generation' and 'iPhone SE 2nd generation'. The price range of each model has also increased over the years.

A-Q8: The most expensive model is iPhone 13 Pro Max *($1099 USD - $1599 USD)*; the more budget-friendly model is iPhone SE 1st generation *($399 USD - $499 USD).*

A-Q9: Despite the increasing price points, the sales of iPhones generally have not dropped but rather, increased.

A-Q10: Apple's best competitor in the Global Smartphone Market 2021 was Samsung. From January to September 2021, Samsung had been taking the lead. It was only until September 2021, that Apple surpassed Samsung in the market, which may be due to the release of the iPhone 13 series.

Critical discussion

This section will discuss the **key decisions** in visualizing the graphs.

Interactivity

Tabset Panel from Shiny is used because it arranges and sorts the plots into a more structured form which allows easier navigation and space efficiency. The idea of having a "Summary" tab for each plot is to have an evaluation of the plot and for users to better understand the data. Shiny also adds more flexibility to the plots as users can switch to view different plots.

ggplotly() and *plot_ly()* from Plotly are used because they add more options to how users can interact with the plots – hovering, zooming, selecting a specific section to view, etc. Overall, when planning visualization, the goal is to make it simple and straightforward to understand.

1. **When plotting Apple's Revenue by Year,** a line graph is used because it is suitable to show time series data – tracking the change of revenue over the years. The dots are added to **emphasize** the exact revenue generated each year. Position is used to encode Revenue to show its **quantitative** value, and Year to show its **order**.

2. **When plotting Apple's Revenue by Quarter** and **by Region,** a stacked/grouped bar chart is used because it is suitable to compare values between different groups – **a part-to-whole relationship.** Also, the grouped bar chart is used particularly because it shows the revenue of individual quarters/regions more clearly side-by-side.

Colour is used to encode nominal Quarter and Region as it is **associative**, where bars of the same colour are perceived as the same Quarter or Region group despite the difference in values. Also, **the Law of Proximity** is applied to group quarters and regions by year instead of the other way round. If we had grouped years by quarter, the graph would've looked like **Appendix 10**, which isn't straightforward enough to understand.

3. **When showing the locations of Apple stores around the world**, a map is used because it is suitable to visualize geographical information. Then, since each location – or region – is already confined to its geographical size and position, colour is used to encode the number of stores for each region to show the quantitative values.

4. **When plotting the Sales and Revenue of Apple by Category**, instead of having all categories plotted in one plot, the plot is **subdivided** by category to avoid over-encoding.

This is because, for each category, there are already 3 variables to be plotted: **Sales, Revenue, Year**; putting all categories in 1 plot will make the plot hypervariable and hard to understand. An example without subdividing is **Appendix 11.**

That said, a bar-line graph with a dual y-axis is used for **each** category. Bar and line are suitable to show **quantitative values** of Revenue and Sales. They're also **distinctive in appearance**, so Revenue and Sales can be **easily distinguished from the graph.**

A dual y-axis is used because we want to **emphasize the quantitative values of both Revenue and Sales.** If we had used only 1 y-axis, the value of another variable would've been unclear, like **Appendix 12.**

5. **When plotting the Revenue Contribution(%) by product categories,** an area graph is used because it shows the **magnitude of different categories over time**, in terms of revenue.

The keyword here is "*over time*". If we had wanted to view the data for a particular year, then a pie chart would've been more suitable. However, in our case, we want to look at the overall winner for revenue contribution and not just the winner for a particular year.

Also, since the revenue contribution will always add up to 100%, an area chart can **clearly show the portion** each category takes – using size as the encoding.

6. **When plotting the Revenue Contribution(%) of different regions to Apple,** the data is **subdivided** by year and 1 pie chart is used to visualize the data for 1 year.

An area graph isn't used because it is similar to the stacked bar chart that shows the actual revenue ($billion USD) generated by regions over time, where the magnitude of each region can already be observed. So for this, we want it to be year-specific.

7. **When plotting the price evolution of iPhones over the years,** a point-range graph is used because it is suitable to show the **quantitative price range of iPhones**, using the start and end prices as the upper and lower limits.

8. **When plotting the relationship between iPhone Sales and Price**, a scatter plot is used because it is suitable to show the **relationship between 2 quantitative variables** by using position as the encoding.

9. **When plotting Apple's and other brands' share(%) in Global Smartphone Market 2021,** a multi-line graph is used because it is suitable to **compare quantitative values** – share(%) – of **different groups over time**. The pie chart is used to show the share(%) of each month for "details on demand".

Reflection

This section discusses what I have learned during the development process.

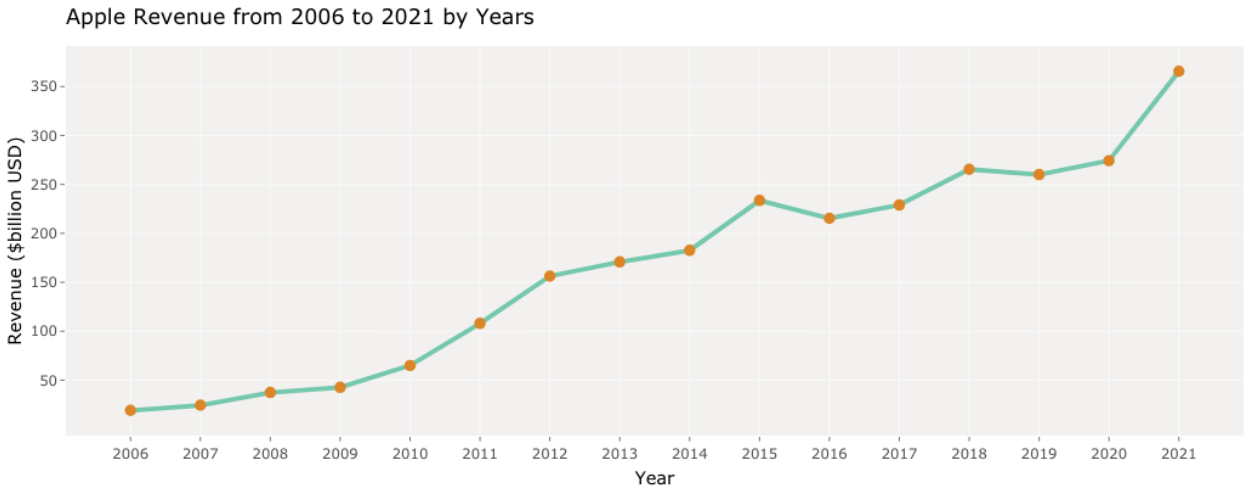
- While plotting has been the main task for this coursework, arranging all the plots in a meaningful way is also crucial. It influences users' experience and thought process as they view the information. The plots should be arranged with a flow, based on priorities.
- For instance, plots about Apple's revenue and sales are arranged to the front in sequence, followed by plots of other information.

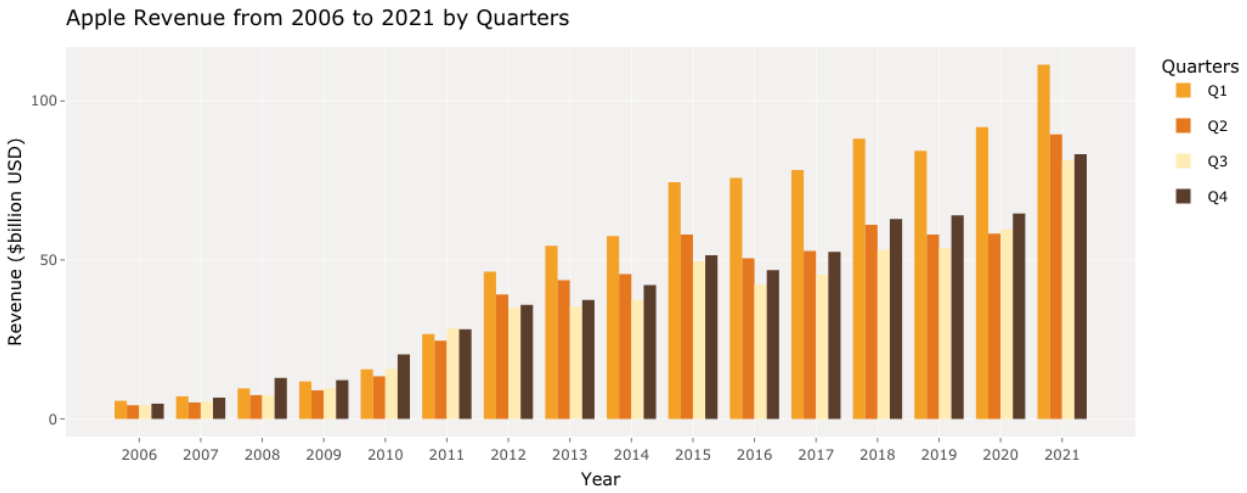
- Besides having a consistent colour scheme, choosing the right colours is also important as it ensures the readability of the plots.
- For instance, when encoding discrete data with colours, colours that are distinctive enough from one another are chosen instead of colours from the same hue but in different tones – dark blue, mid blue, and light blue.

- Adding interactivity to the work is beneficial. It doesn't only help in saving space but also makes the information more organized. Interactivity also introduces more flexibility to the plots as users can apply parameters to the plots.

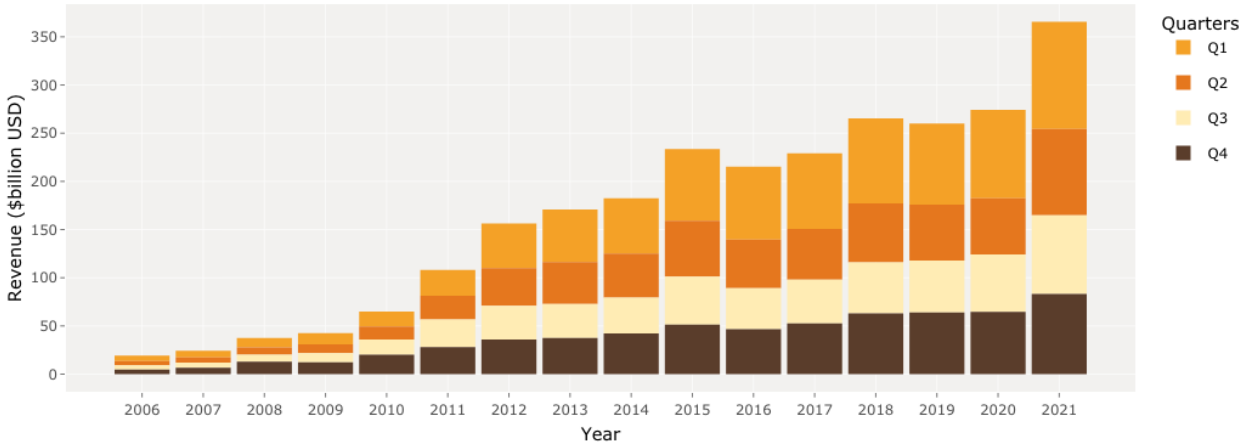
Appendix

1. Visualizing Apple's Yearly and Quarterly Revenue from 2006 to 2021

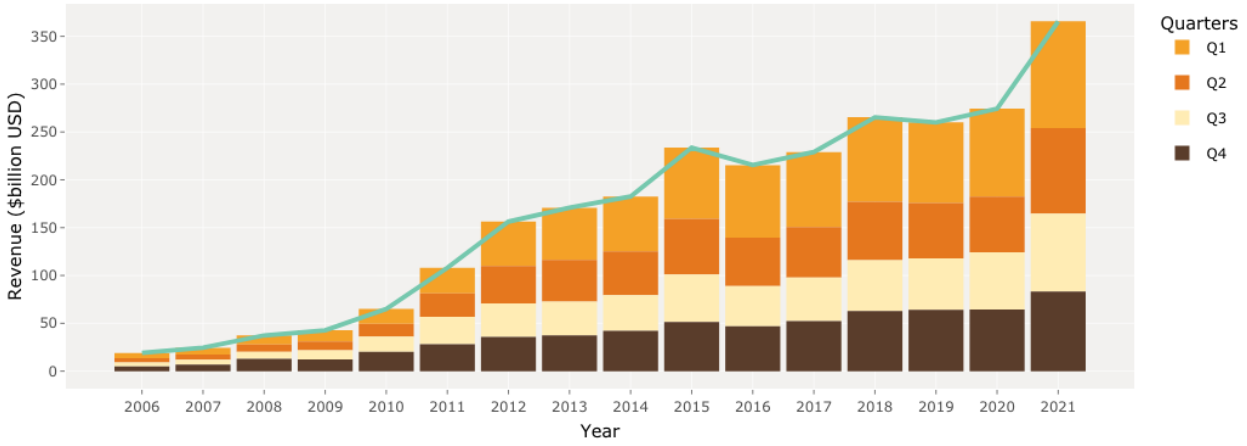




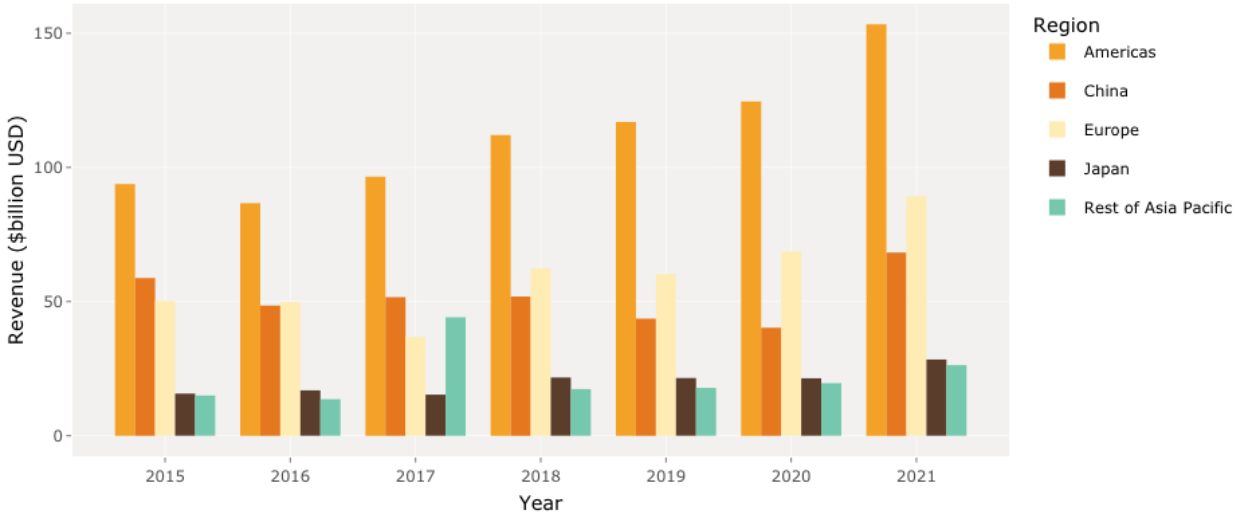
Apple Revenue from 2006 to 2021 by Quarters

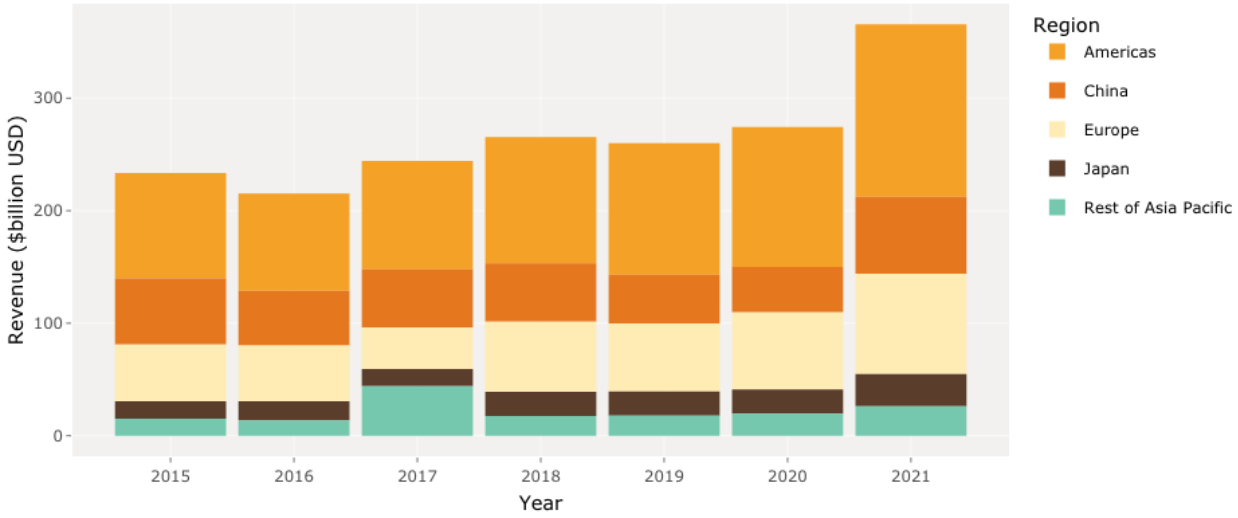


Apple Yearly Revenue from 2006 to 2021 by Years and Quarters

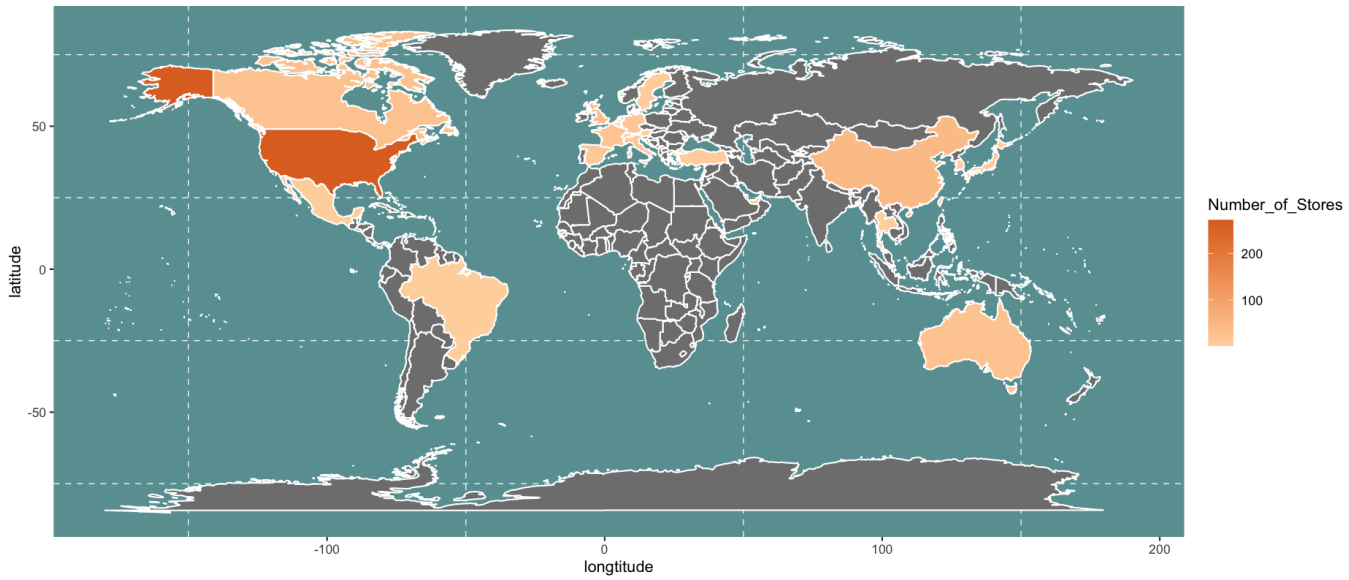


2. Visualizing Apple's Revenues by Region.

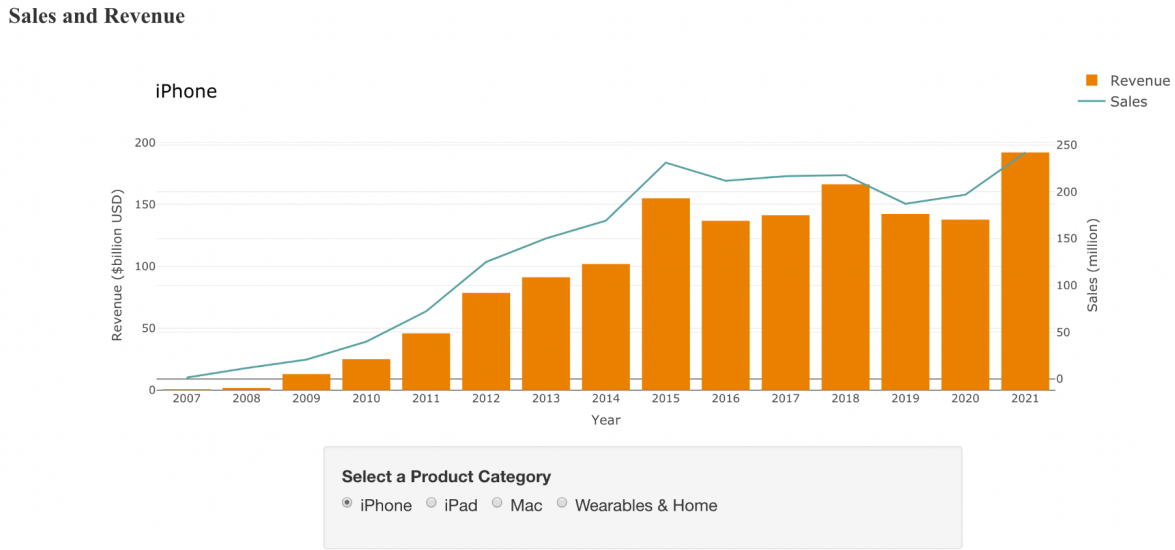




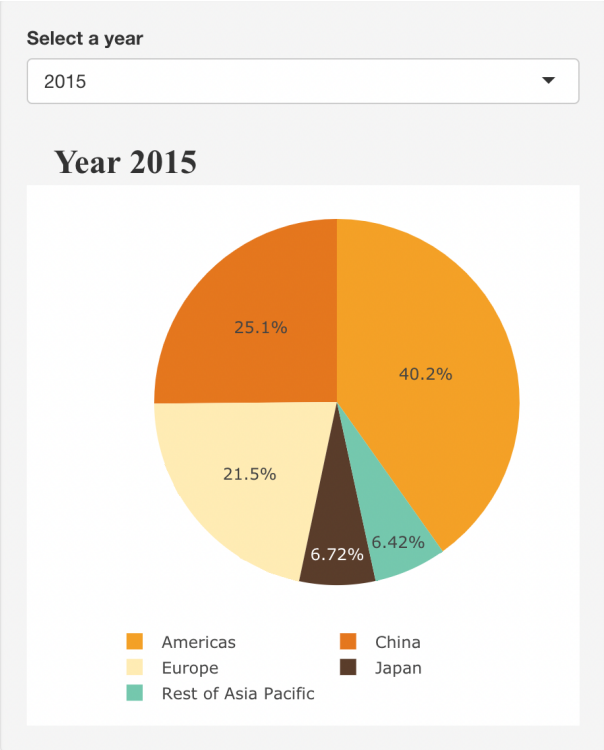
3. Visualizing Apple's stores worldwide

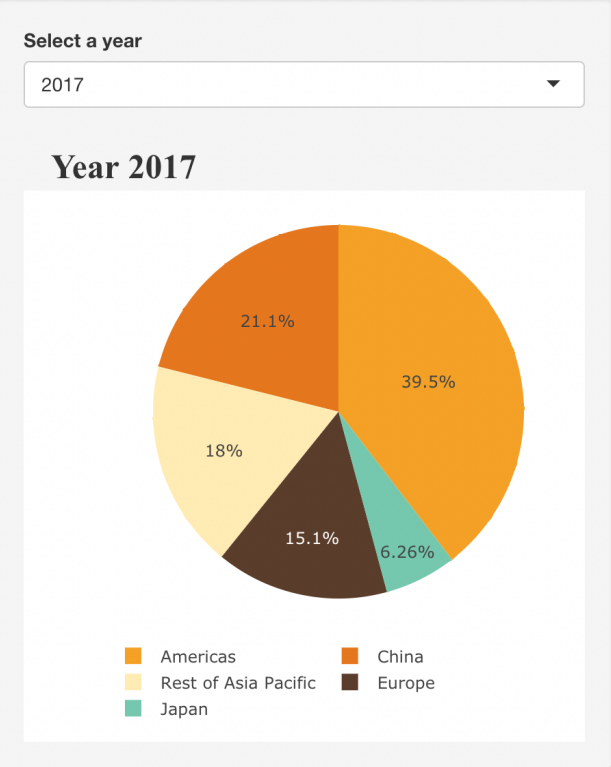


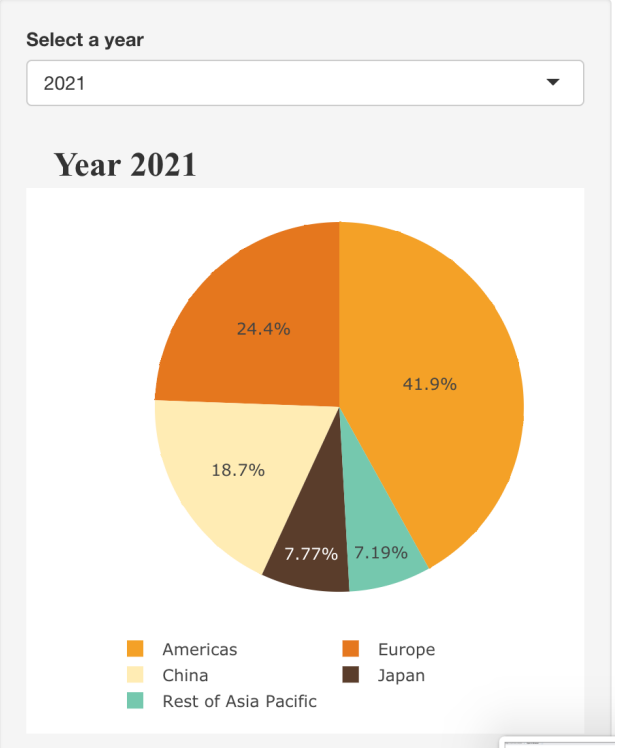
4. Visualizing the sales and revenues of different product categories.



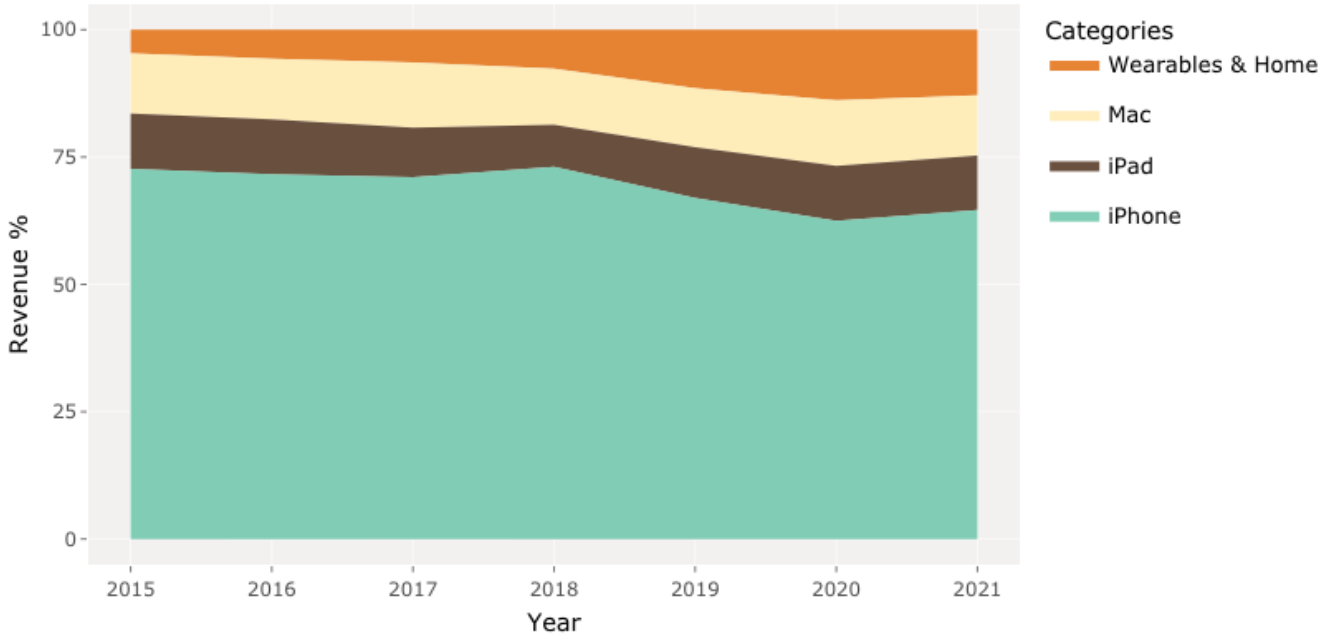
5. Visualizing Revenue Contribution (%) by Region for each year from 2015 to 2021.



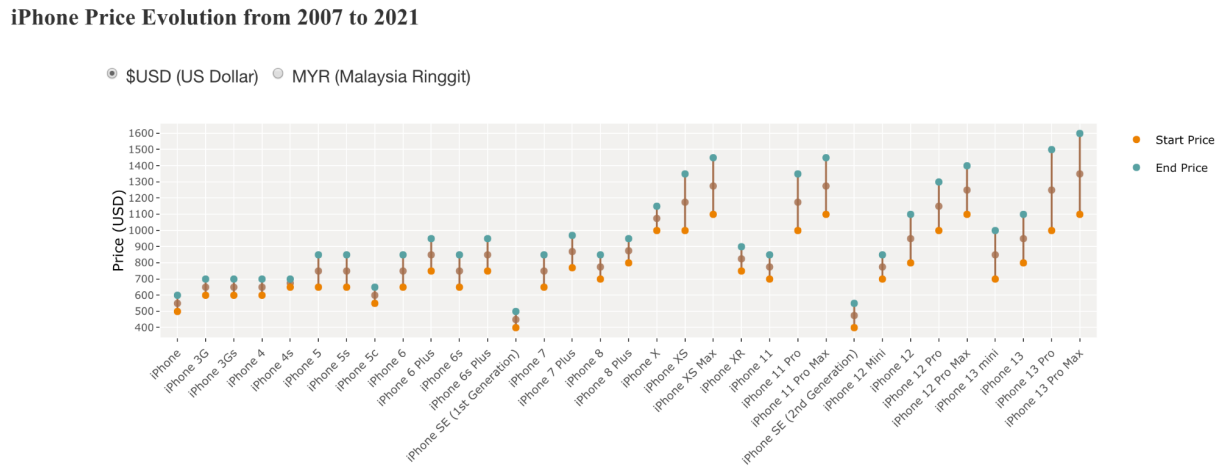




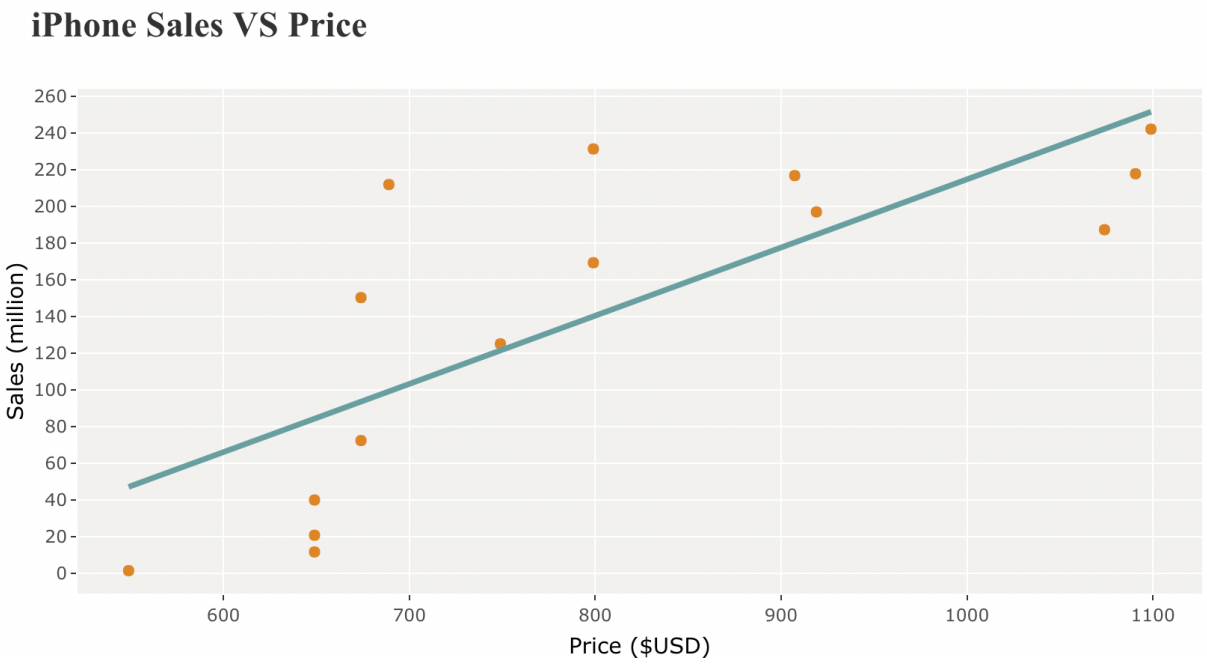
6. Visualizing the Revenue Contribution of each product category to Apple's revenue from 2015 to 2021



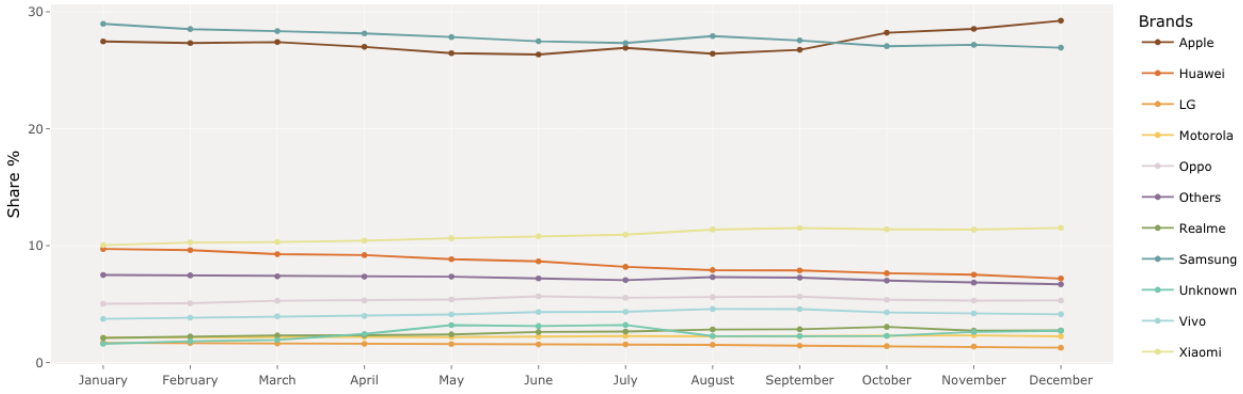
7. Visualizing the price evolution of iPhone



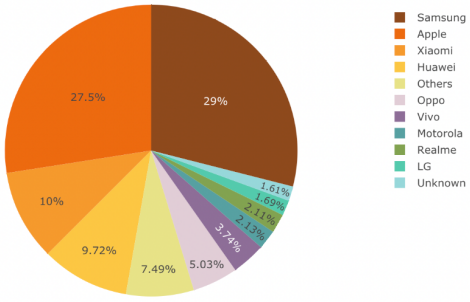
8. Visualizing the relationship between iPhone Sales and Price



9. Visualizing the share (%) of Apple with other brands in the Global Smartphone Market 2021.



Global Market: January 2021

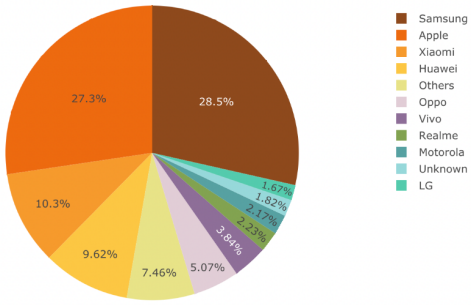


To view the Market Share of a particular month

Select a Month to view

January

Global Market: February 2021

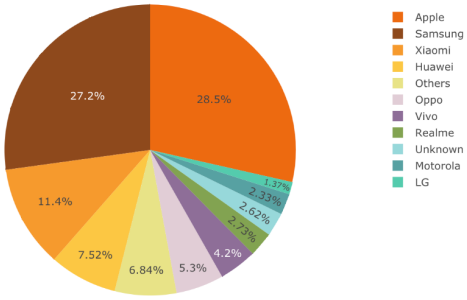


To view the Market Share of a particular month

Select a Month to view

February

Global Market: November 2021

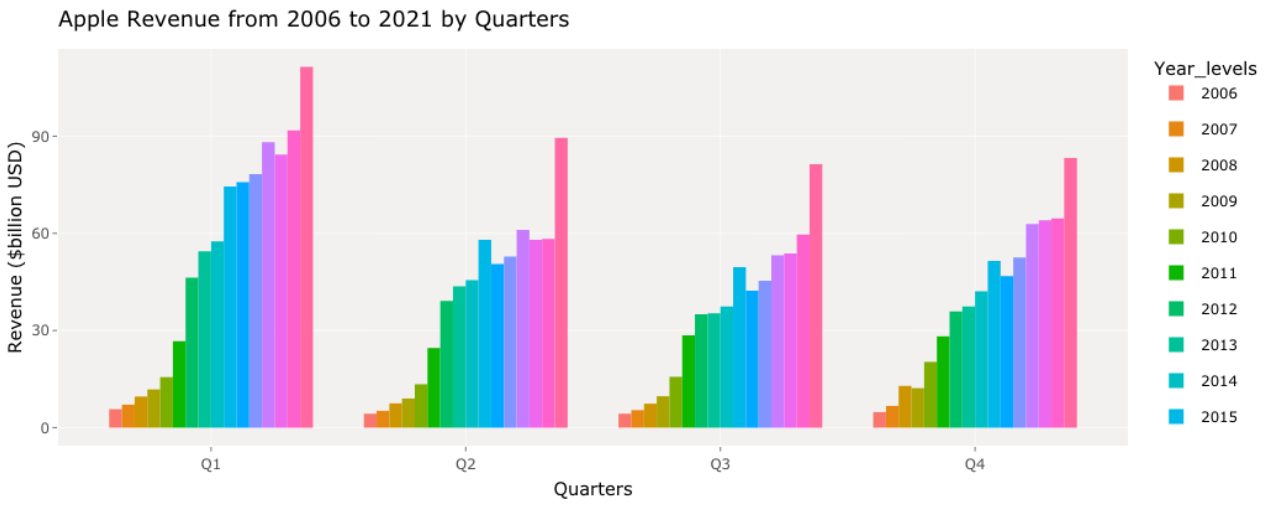


To view the Market Share of a particular month

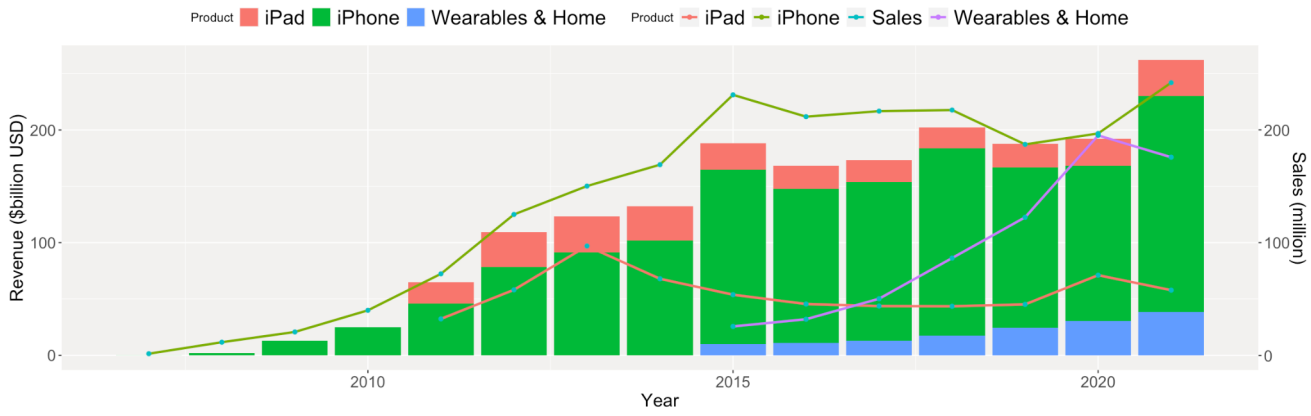
Select a Month to view

November

10. Group Years by Quarters instead of Quarters by Years



11. Visualizing the Sales and Revenues of all product categories in 1 plot, without subdividing.



12. Visualizing the Sales and Revenues of a product category without using a dual y-axis.