

UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE  
INGENIEROS INFORMÁTICOS



Máster Universitario en Ciencia de Datos

# HOMEWORK 1.2: DATA MANIPULATION AND VISUALIZATION

EL ABASSI WIDAD

JIAYI LIN

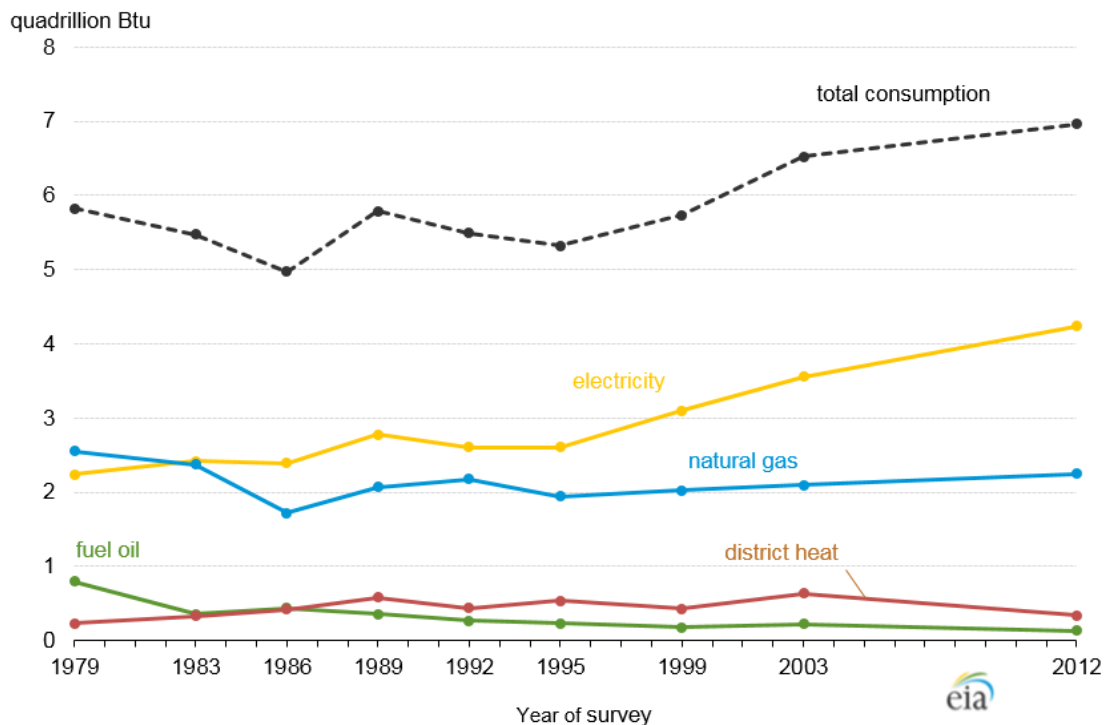
XIAO LUO

AZIZ NABIL

Madrid, November 3<sup>rd</sup> , 2019

## ✓ Case 1:

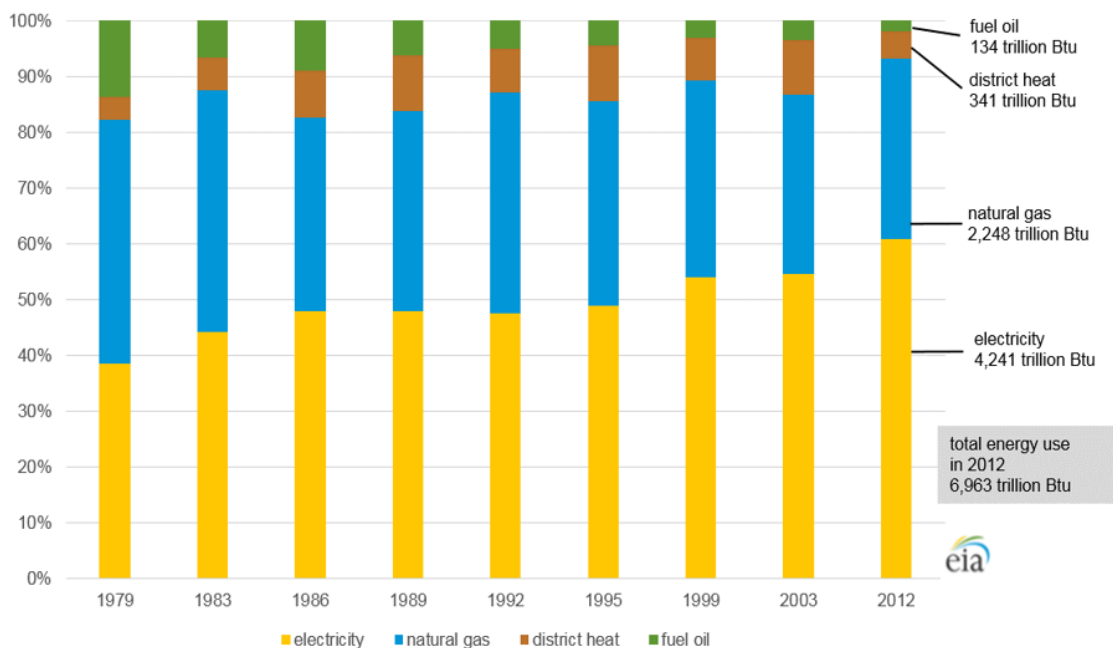
Figure 1. Total electricity usage has increased more than the other energy sources since 2003



Source: U.S. Energy Information Administration, Commercial Buildings Energy Consumption Survey.

**Review:** This line chart pretends to show the increasing of total electricity usage among the other energy sources. The type of the chart is correctly applied, because it shows the trend of energy consumption over a time series. By using different colors for energy sources, it makes the presentation clear and simple. The accuracy in this graph is demonstrated by the points on each line, although it can be improved even more by adding specific values on the interested points. The way of distinguishing types of energy resources with each color can be changed by creating a separated chart legend to improve the clarity.

**Figure 2. Electricity now accounts for 61% of all energy consumed in commercial buildings**



Source: U.S. Energy Information Administration, Commercial Buildings Energy Consumption Survey.

**Review:** This stacked bar chart shows the evolution of the proportion of different energy consuming in commercial buildings, so we can say that the choice of plot was good. The use of different colors gives the viewer more clarity and ability to recognize the consumption of different type of energy, the space between the bars makes it easy to distinguish the evolution by years. The graph is simple to read because of the Claire legend and title and the organized struct. We can say that the graph makes it easy to study the evolution of the consumption of the electricity regarding other types of energy, but the limit for this graph is that we can't study the evolution of the consumption of each type at once due to the lack of numerical values, the second limit is that the figure doesn't take in consideration all the years the gap between years is big 3 or 4 years. The other note is that the scale of years is not constant, those lacks make it difficult to find patterns for the evolution. The accuracy in the graph can be improved by showing exact value on each color of the bar. Further detail information on the right side can be deleted to make the graph simpler because the objective of this bar chart is to show the proportion instead of quantity consumed.

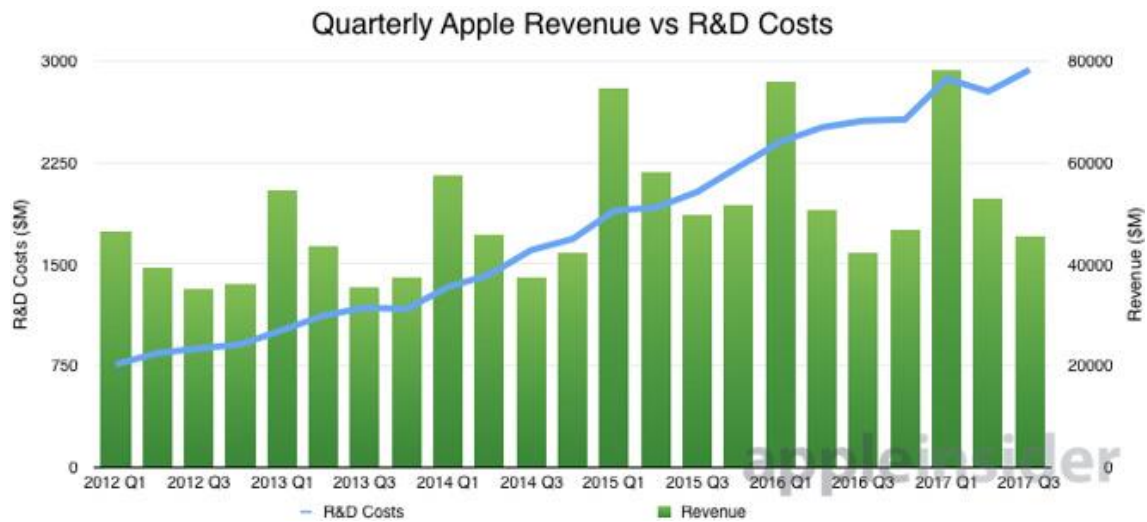
**Comparison:**

Both charts shows clearly the goal of each case, one to demonstrate the trends of each energy consumption along the year by using line chart, the other on shows the proportion of each energy consumption in each year. But even both graphs gives possibility to study the evolution of the consumption over time but there's some main differences such as the possibility to fit data into groups or not , the ability to get a specific value in time and analyzing values in continuous way, the spread of data across different values.

**Conclusion:**

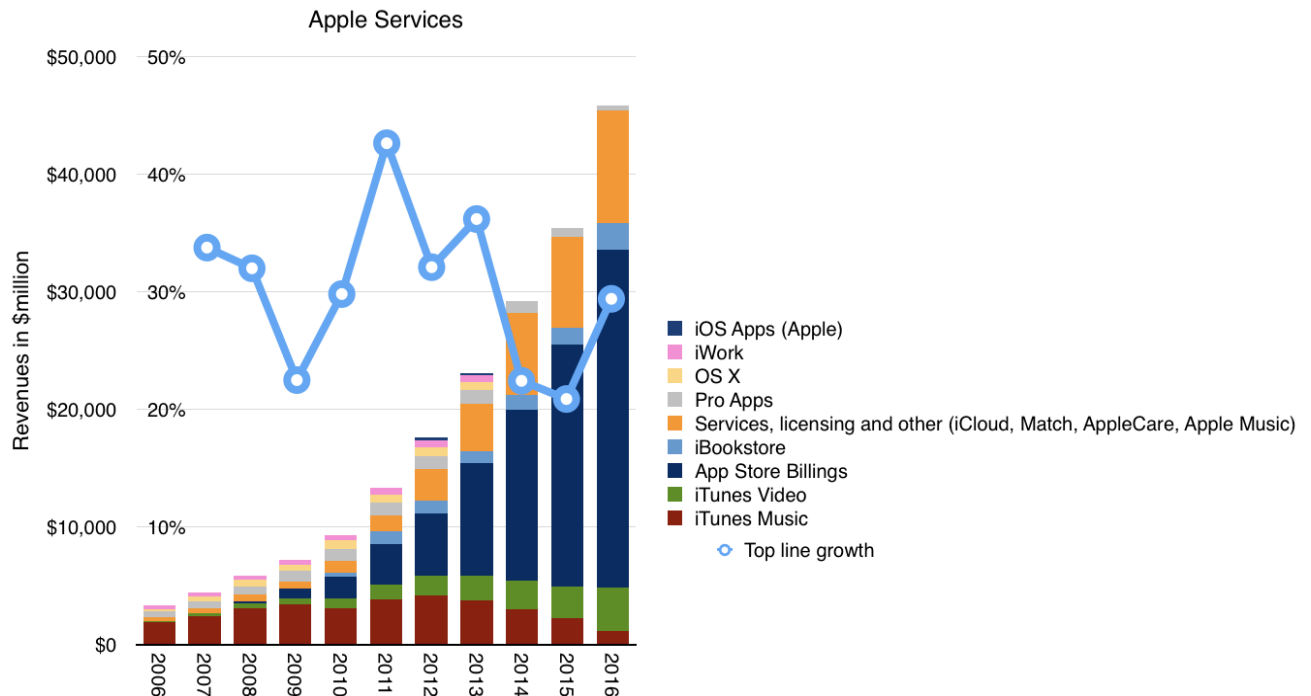
There's a lot of main point that are important in order to analyze the power, the accuracy and the clarity of the visualization of data on graphs such as colors, structure, the choice of the right chart ... etc. in our analysis to graphs we conclude that Bar graphs can show trends over time as groups (as in Figure 2), but line graphs have an advantage in that it's easier to see small changes on line graphs than bar graphs, and that the line makes the overall trends very clear. They are less versatile than bar graphs, but better for accuracy.

✓ Case 2:



**Review:** The figure shows the evolution of Apple's quarterly revenue with bar chart and R&D costs with line chart, both good choices for represent numerical data on timeline. The former's scale is shown at left and the latter at right. Both variables are represented in the figure not only with different type of plot, but with different colors, which brings high level of clarity, making it easier for possible viewers. It also contains the minimum amount of supportive texts, which makes the figure very simple.

Furthermore, the number of labels does not match the number of bars and it's not even a multiplier of it (23 bars with 12 labels) it's very confusing for the reader and difficult to understand, also the labels are not separating the groups, they're all concatenated with each other's. A possible improvement to the figure could be separating it into 2 figures, since mixing different scales at left and right of plot isn't a recommended practice.



Staying in the same field we have this chart published by Asymco in their post about orthogonal Pivots faced by industries. This chart is used in order to help illustrate the business growth of Apple revenue with highlighting the services that contributed in this.

This statistic graph is a time series Line Chart overlay on a Bar Chart. On one hand, categorical information were encoded in colors to distinguish groups and also highlights this specific part of the dataset ( show the magnitude of each service on the revenue growth ), on the other hand, a line chart was used to demonstrate the yearly basis growth for this company starting from 2006.

### Comments about Representation:

- ❖ The legend is disconnected from the graph and I think this is maybe due to the number of items in it. Also, the legend order matches the the bars.
- ❖ The use of colors made it easier to perceive the difference in the contribution of services over the years.
- ❖ The use of stacked bar chart in this scenario instead of grouped bar chart is very clever in order to avoid o more large chart.

### Possible Improvements:

- ❖ Title isn't descriptive: This is a combined chart and all of the pertinent information must be contained in the title, while it's not the case here, only a part of the description is mentioned above.
- ❖ The Order of the items (services) in bars must follow a chronological order ( from the old ones to the recently developed ), so I think it could be best if we put the most recent services at the top of bars.

- ❖ This a combined chart with two datasets that contains data with different measurement which result in a dual axis chart, except that here, we note that the same axe was used to represent two different data series with different scales ( line growth is % and services revenue in \$millions).
- ❖ No source of the chart was mentioned below.

### **Comparison:**

Both combined charts needs improvements: in the first chart, we believe the choice of a side by side chart could've been more appropriate, and if we take into account the number of bars we must have chosen a vertical bar chart instead of the horizontal one, as for the labeling, we should consider splitting the labels in 2 lines ( first for the date and the second for the quarter ) because the way they're represented makes it difficult to understand which label correspond to which bars ( the goal of labels is to separate the groups which was not achieved in this figure ). In the second figure, the choice of stacked chart type was a reasonable one, seeing that we have more than 10 bars with short labels. However, the title was poor and the services was unreasonably stacked in the bars.

### **Conclusion:**

Normally, the combined charts are actually harder to read, because the view is overloaded with columns, lines, colors and sometimes the scales of dual axis charts are arbitrary and can therefore mislead readers about the relationship between the two data series (which is the case for the first figure). So, we must be cautious of which type of chart must be applied to data because the user not only might be confused by the information, but, more importantly, could make bad decisions based on such a presentation.

### **Sources :**

**Chart 1, 2 :** <https://www.eia.gov/consumption/commercial/reports/2012/energyusage/>

**Chart 3 :** <https://appleinsider.com/articles/17/08/09/though-apples-rd-spending-is-massive-its-still-more-efficient-than-all-other-competitors>

**Chart 4 :** <http://www.asymco.com/2017/10/04/orthogonal-pivots/>