



# Assignment 3

## DATA VISUALISATION

Arnav Malhotra | 17317424 | 10-03-2019

# Visualisation Analysis

## Farr (Late 1840s):

The concerned dataset was visualized by William Farr depicting Temperature and Mortality of London (For every week of 11 years, 1840-50).

It showed **static** data in a **quantitative** fashion having **absolute measurements**. The **visual encoding channels** used are **colour**, **size** and **position**. the distance between each concentric circle represents either 100 deaths or 10° of temperature.

The purpose of this diagram was to show a correlation between temperature and mortality rate of London. More specifically, it supported the miasma theory of cholera, meaning that cholera was airborne. People living at lower elevations in London (where temperature was also lower) were more susceptible of cholera.



Figure 2: Comparison between the pattern of patents from Apple (Left) and Google (Right)

These places also had River Thames in common. Later, physician John Snow disproved Farr's theory of miasma by concluding from his own study that cholera was waterborne.

Although Farr's visualization is very intricate and has a huge amount of information, that does not necessarily mean that it is a better visualization. By including so much information in such a small space, it becomes rather confusing and fails to accomplish tasks such as **clustering**, **location** and **association**.

## Wilson (2017):

This dataset shows the difference in the way how Apple and Google operate. Each blob is a patent inventor and as many patents have multiple inventors, lines are used to link inventor and co-inventor.

It is shown in a **categorical** fashion. Further, the values are **ranked** with smaller blobs representing lesser patents by an inventor and larger blobs representing more patents by an inventor. Darkness of blobs represents the level of authority/designation of the inventor, i.e., the darker the blob, the higher the position of the inventor. However, Green blobs represent the founder/co-founders of the company. The inventors at the peripheries represent ones whose patents have been acquired with them not being a company employee. Therefore, the **visual encoding** channels used would be **size** (of blobs), **colour**, **position** and **brightness**.

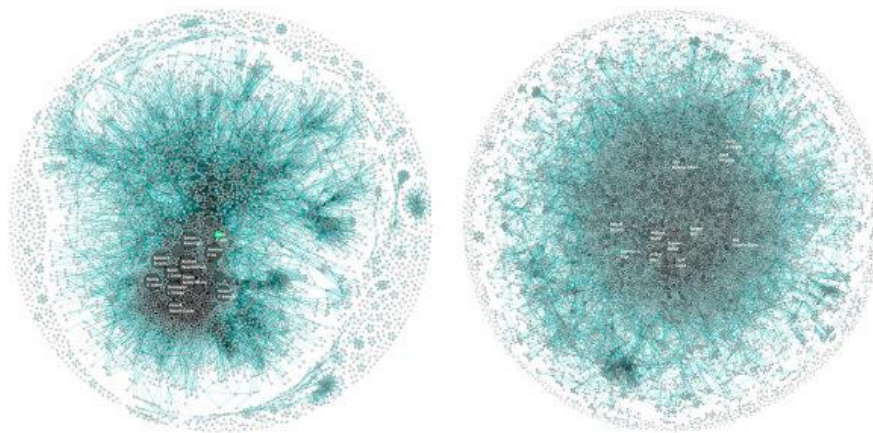


Figure 2: Comparison between the pattern of patents from Apple (Left) and Google (Right)

By looking at the above diagram, the companies' development structures can be discerned. In case of Apple, it seems to have a centralized system as compared to that of Google. "Over the past 10 years Apple has produced 10,975 patents with a team of 5,232 inventors, and Google has produced 12,386 with a team of 8,888," writes Wes Bernegger, data explorer at Periscopic. The graph also cements the fact that in case of Apple, most of the patents are made by their design team led by Jonathan Ive. On the other hand, Google seems to have a more democratic approach with a larger number of independent people and lot of patents from open-source community.

## Yau (2019):

This shows us visualization of data from 2015-2017 release of the National Survey of Family Growth, which is run by the Centers for Disease Control and Prevention.

The data comprises of various timelines of married people focusing on at what age and how many children they have. However, for the interactive animation, only 1000 timelines are simulated. Though it looks like the data is in ranked categorical fashion due to the different sizes of the green circles, yet the datatype used is **discrete quantitative** type as those circles represent whole numbers.

The black dots which travel as the animation progresses each represent a mother. The green circles represent mothers of that particular age (denoted by horizontal axis) and having a number of children (denoted by the vertical axis). The size of the green circles denotes the number of mothers belonging to that category. The visual encoding channels used are:

- 1) **Position**
- 2) **Size**
- 3) **Colour**

Looking at the animation, we can see that:

- 1) The births are concentrated between ages 17 to 30 and 1 to 3 children.
- 2) At the start, the number of people having more than 2 children was higher but as time progresses, comparatively more people want to have lesser number of children.
- 3) The number of firstborn children in case of 30s rises as the time progresses.

# Visualisation Design

## About the Dataset:

The dataset being used here is **Sexual offences in England and Wales: year ending March 2017**. It has been taken from **Crime Survey for England and Wales**.

Mainly, it contains data about-

- 1) Prevalence of Sexual Assault among adults aged 16 to 59, by type and sex.
- 2) Estimated numbers of victims of sexual assault since age of 16 by type, sex and year.
- 3) Number and rate of sexual offences recorded by the police.
- 4) Victim-offender relationship by sex and year.
- 5) Influence of alcohol and drugs in rape or assault.
- 6) Outcome of police investigation.
- 7) Effects of rape or assault on victims and impact on their work.

But **Table 8: Police recorded sexual offences, by offence, year ending March 2003 to year ending March 2017 and percentage change between year ending March 2016 and year ending March 2017** is of particular interest.

The attribute type is **quantitative** with **measurements**. The dataset type is a **list** of **static data**.

## Visualisation Tasks:

Identification, Categorisation, Association, Comparison, Ranking and Correlation are the visualisation tasks which need to be taken care of. Looking at the dataset, we can compare the rate of rape vs other sexual offences over the years and also in terms of sex.

## Encoding Channels:

Line graphs, Scatter plots, or Tree map charts might be fit for this visualisation. Various encoding channels can be used such as **colour** to represent sexes, **size** and **position** would represent the number of the concerned parameter.