

DATAVIS

VISUALISATION OF CLASSIFICATION MODELS.

EXPLORATORY DATA ANALYSIS (EDA)

- ▶ In statistics, exploratory data analysis (EDA) is an approach to analysing data sets to summarise their main characteristics, often with visual methods.
- ▶ A statistical model can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modelling or hypothesis testing task.

EDA



Why do EDA?

EDA

- ▶ Some of the main reasons:
 - ▶ Detecting Mistakes.
 - ▶ Checking Assumptions.
 - ▶ Preliminary Model Selection.
 - ▶ Determining relationships among explanatory variables.
 - ▶ Assessing direction and magnitude of relationships between variables.

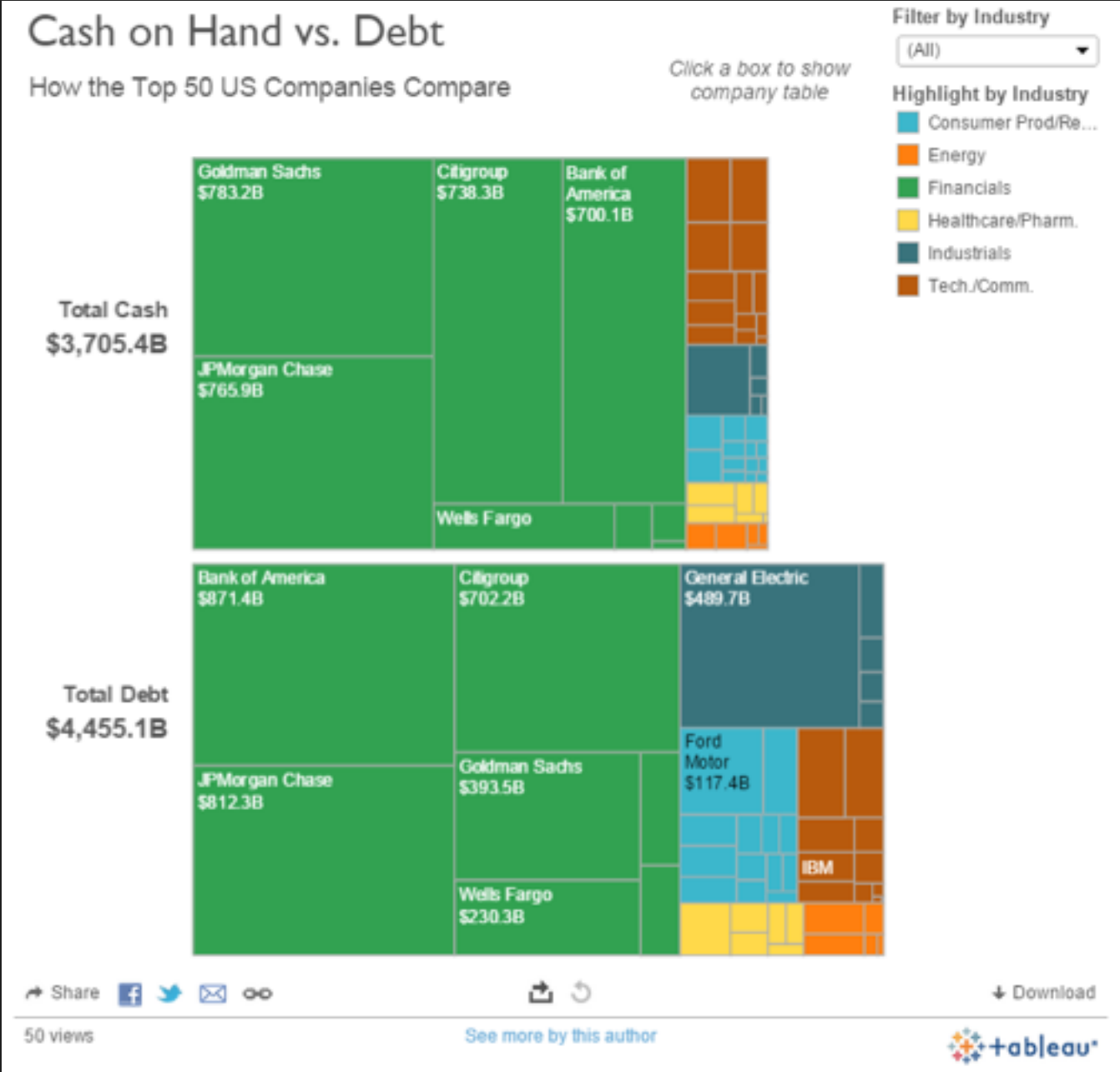
VISUALISATION TOOLS

- ▶ We have so far worked with:
 - ▶ Matplotlib.
 - ▶ Seaborn.
 - ▶ Pandas plotting.
 - ▶ Tableau.
- ▶ And I encourage you to also try: plot.ly

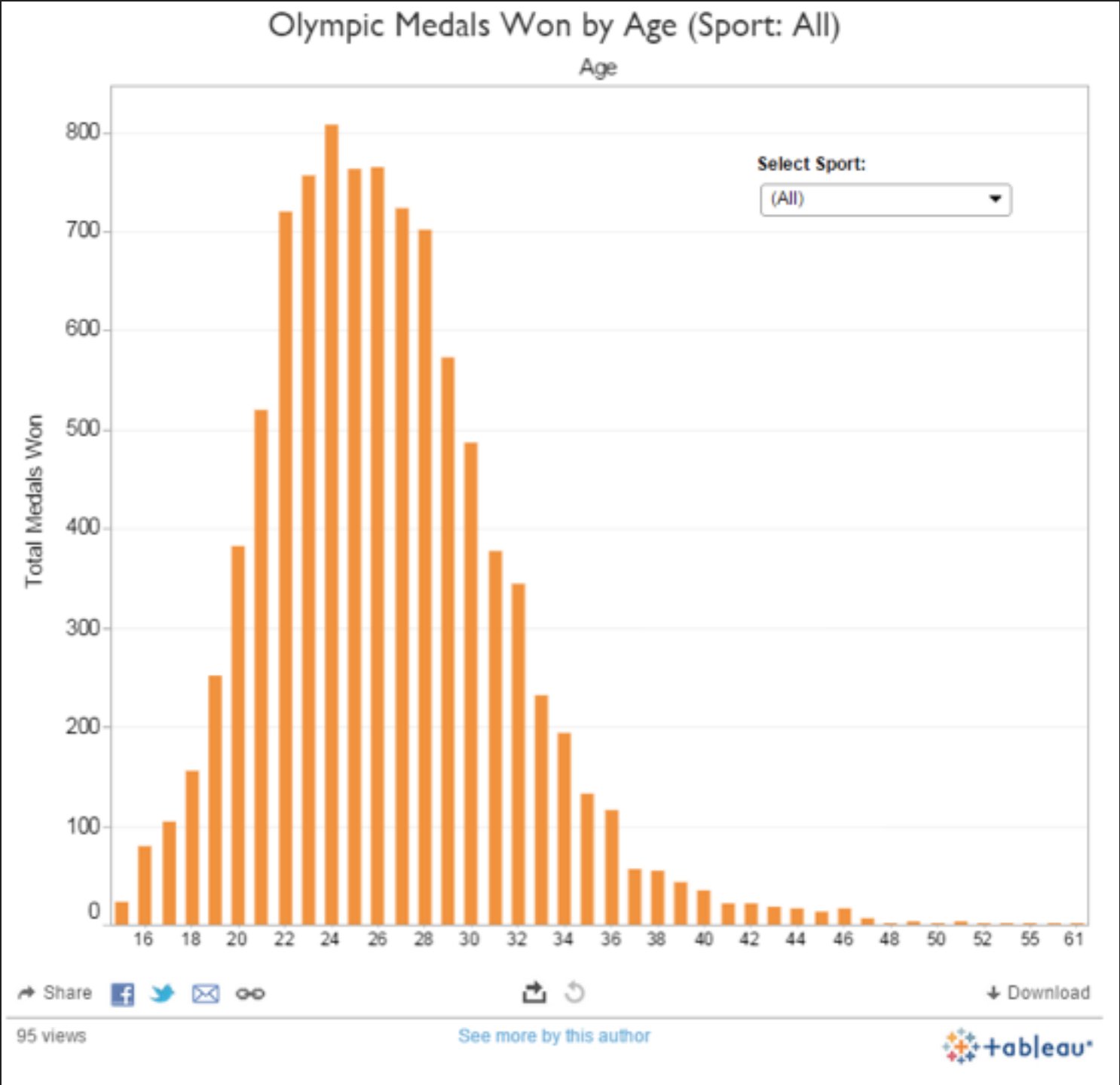
VISUALISATION TOOLS

- ▶ What's your favourite?
- ▶ Pros/Cons of each?

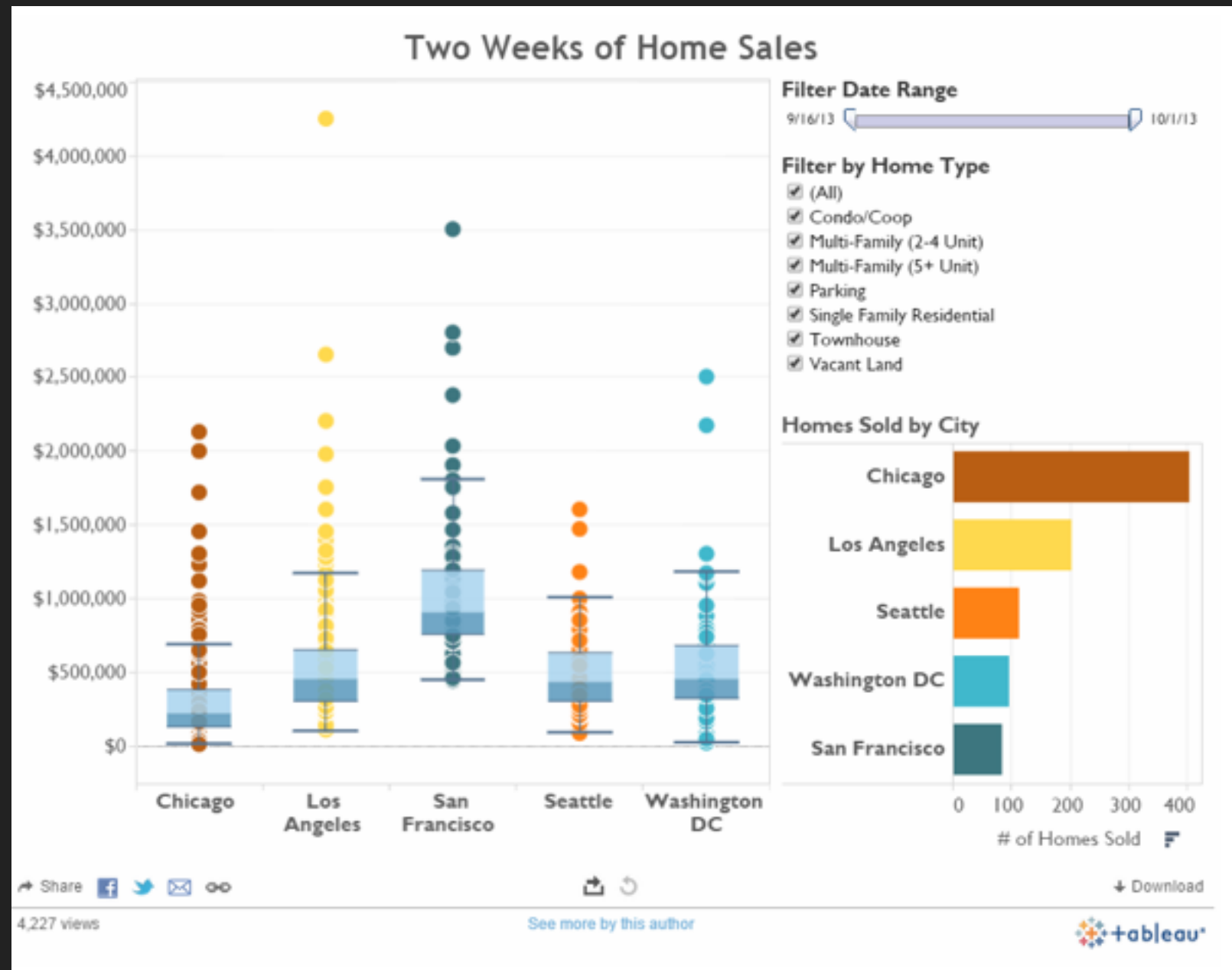
TREEMAP



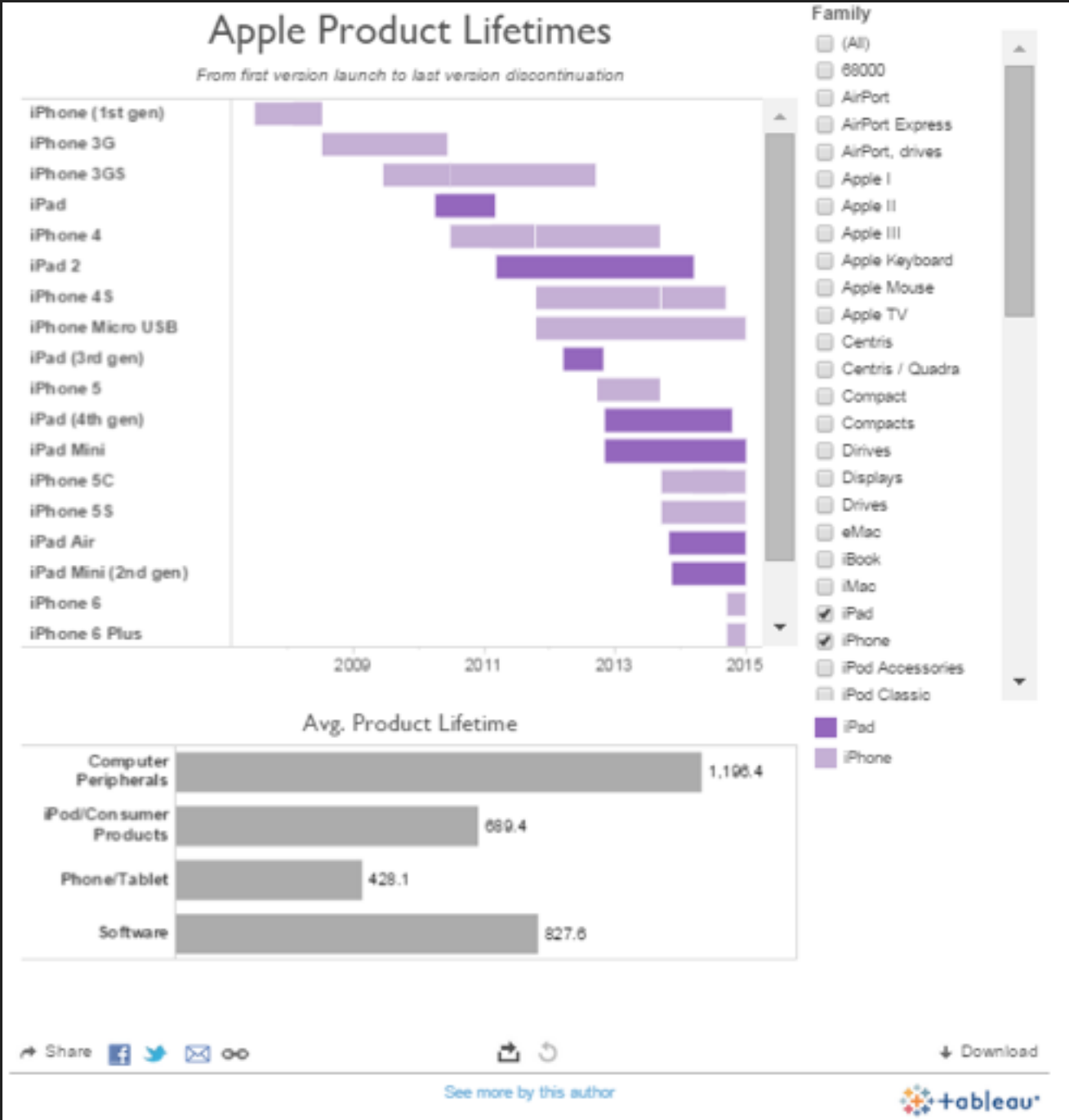
HISTOGRAM



BOXPLOT



GANTT CHART



WORD CLOUD



TABLEAU FOR CLASSIFICATION PROBLEMS

- We are going to use Tableau to explore a breast cancer data set.

