# Jupyter Notebook Mastery – Theory Revision

## Problem solving with data

* Problems should be specific
* Goals are broader, while objectives are specific and executable
* Hypothesis should be testable and falsifiable
* Do not use news articles and blog posts as a reliable source for your research
* Plan your work!

## Introduction to Jupyter Notebook

* Project Jupyter consists of more than just the Jupyter Notebook app
* Jupyter Notebook is not suitable for large scale production applications
* There are three types of cells:
  + Code
  + Markdown
  + Raw NBConvert

## Principles of NumPy

* NumPy arrays are faster than Python lists
* Vectorization substitutes loops
* A NumPy array consists of only one datatype
* We can slice arrays by using a comma-separated list of indices

## Principles of Pandas

* Series is a one-dimensional array-like object
* DataFrame represents tabular data
* Pandas provides flexible ways of handling missing and duplicate values
* We first split and then aggregate our data for the purposes of visualization

## Principles of Matplotlib

* Do not cram your visuals with unnecessary information
* Use up to 4 colors in a visual
* Be careful with the scales in your visual
* Try multiple times until you get it right

## Data storytelling

* Make sure you know the audience
* Be clear of what your goal is
* Hook, intro, tension, climax, resolution
* Format your text properly
* Use Rise as presentation tool