LEARNING LAB, Part 2

Learning Lab, Part 2:

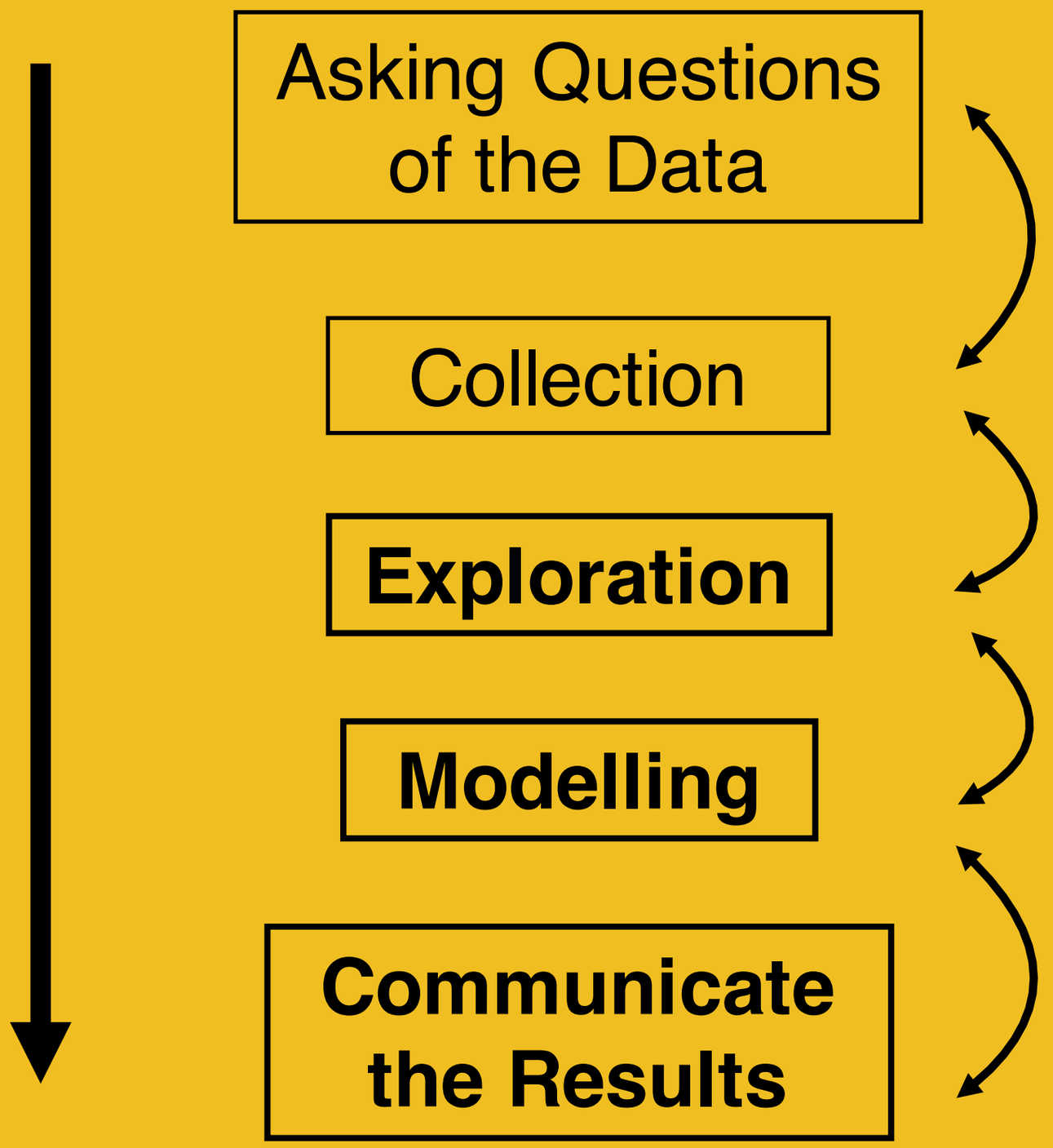
An interactive, guided session where students can follow along on a practical data science project. Learn how to explore, analyse and visualise data using Excel, Tableau and R.

✓ Exploratory Data Analysis

✓ Modelling

✓ Data Visualisation & Communication

The Data Science Process



Part 1

Learning Lab Part 2

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is the process of summarising data using summary statistics and data visualisation

Exploratory Data Analysis

Why do we need to explore the data?

1. To spot problems

> Missing values, bad fields or variables, identifying outliers

> Identify what needs formatting and cleansing

Exploratory Data Analysis

Why do we need to explore the data?

2. To give a sense of the data

> Build a more detailed picture of the data

> Explore summary statistics

> To refine our question (if necessary)

Exploratory Data Analysis

DEMO

Exploratory Data Analysis

Let’s explore our data on diabetes, identify some problems, undertake basic data cleaning and create some quick data visualisations.

Please feel free to follow along or explore for yourself. This demo will use Excel

Modelling

Modelling

Modelling is the stage where we extract valuable insights from the data

Modelling

Common Modelling Tasks:

Classification - Finding out if something belongs to one thing or another

Scoring - Predicting or estimating a numeric value, such as price

Ranking - Ordering items by preferences

Clustering - Grouping items into most-similar groups

Finding Relations - Identifying correlations or causes

Characterisation - Plotting and report-writing

Modelling

Modelling

> For each modelling task, there are multiple options to choose from

> Evaluating models is very important.

Example questions:
Is the model accurate enough, is it generalisable and how does it compare to other models?

Modelling

Please be mindful…

Correlation does not imply causation

Modelling

DEMO

a different question to answer!

Modelling

Let's use predictive analytics to forecast demand for diabetes medication in Leeds. We'll look at choosing the right model and finally visualise the results.

This demo is in R, a powerful open source statistical programming tool that can seem a bit daunting. Feel free to follow along with the process or explore the code for yourself. The code is publicly available.

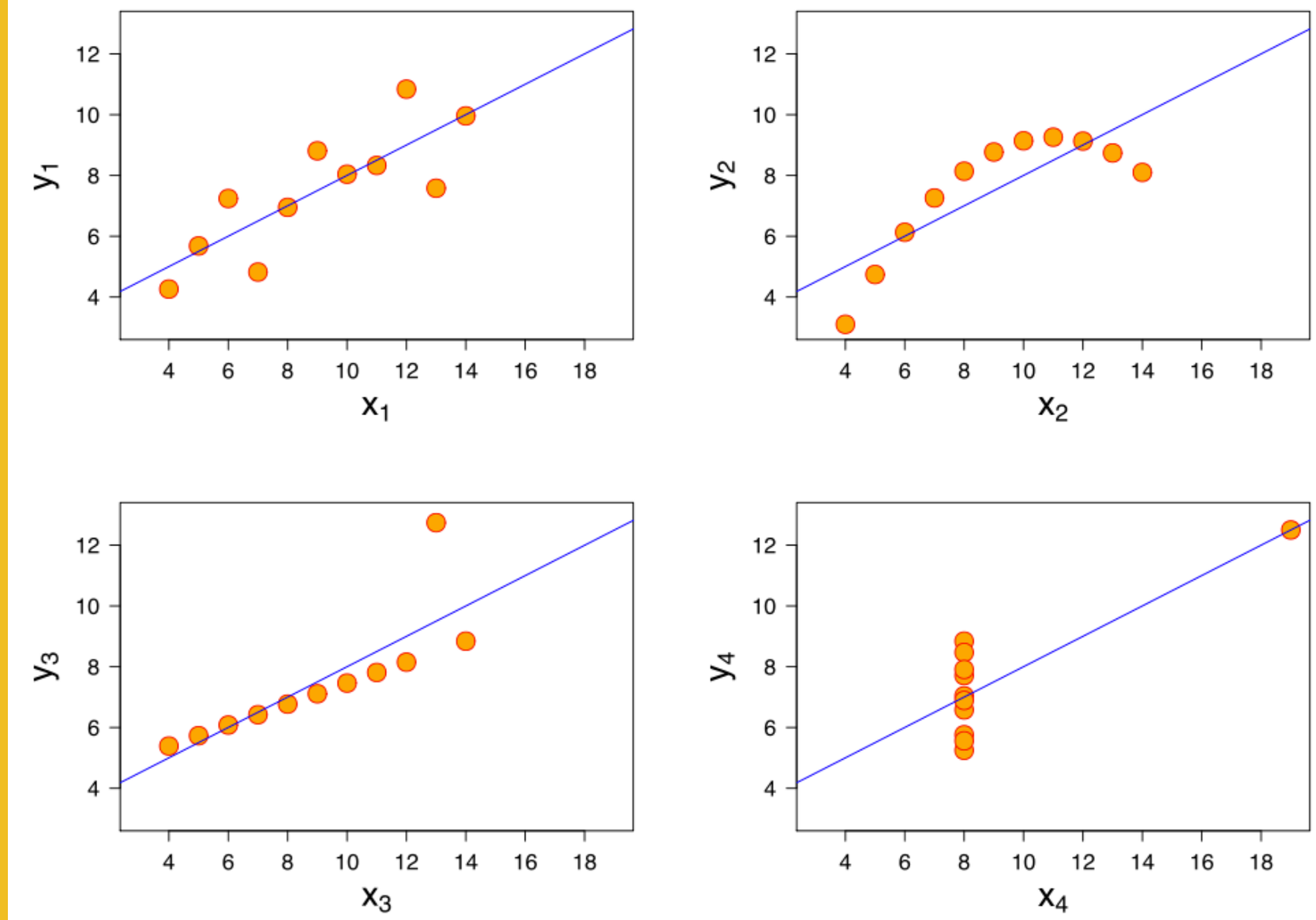
Data Visualisation

The ability to communicate information and the results of a data science project is hugely important.

Storytelling with data is a powerful way to communicate your message

But it should be grounded in some rules

Data Visualisation and Communication



These charts all have the same mean and standard deviation

Source: Alan Smith, FT

Data Visualisation and Communication

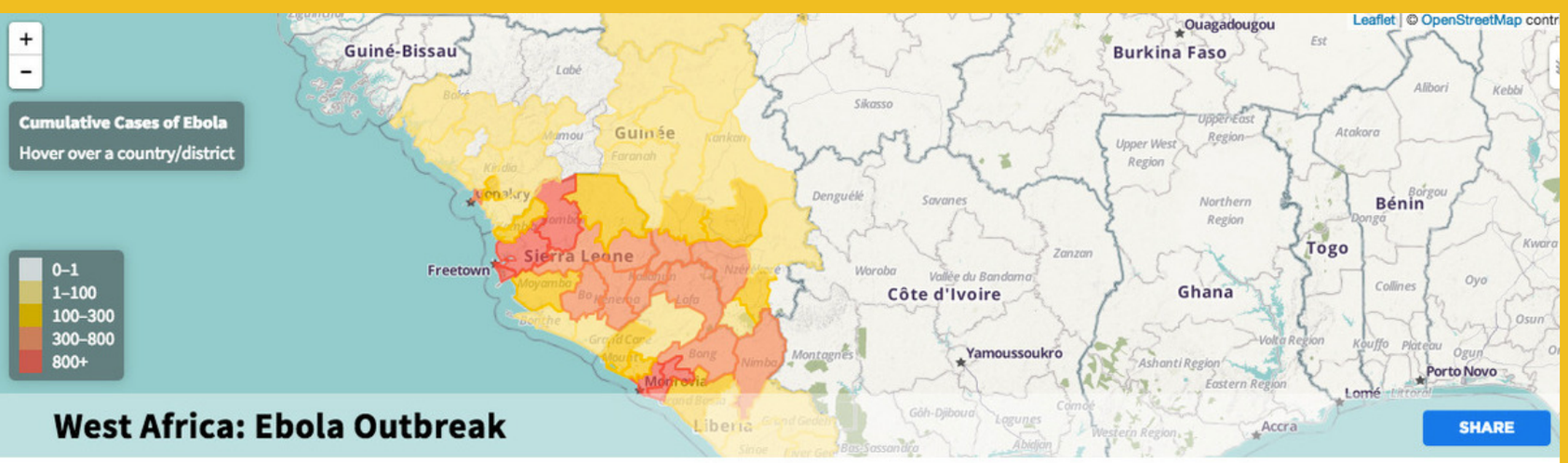
Exploratory Data Visualisation

> The user is free to ask questions of the visualisation and explore and find insights

> Discovery a key element

> No single narrative

> Often interactive



West Africa: Ebola Outbreak

Cumulative Cases of Ebola
23,948
WHO
Data - Explore - Mar 02, 2015

Cumulative Deaths from Ebola
9,729
WHO
Data - Explore - Mar 02, 2015

Response Plan Coverage
57.3 %
OCHA FTS
Data - Explore - Feb 18, 2015

People Receiving Food Assistance
3.3 million

Open Ebola Treatment Centers
49

Currently Affected Countries
4

Humanitarian Data Exchange (HDX)
www.data.hdx.rwlabs.org/ebola

Data Visualisation and Communication

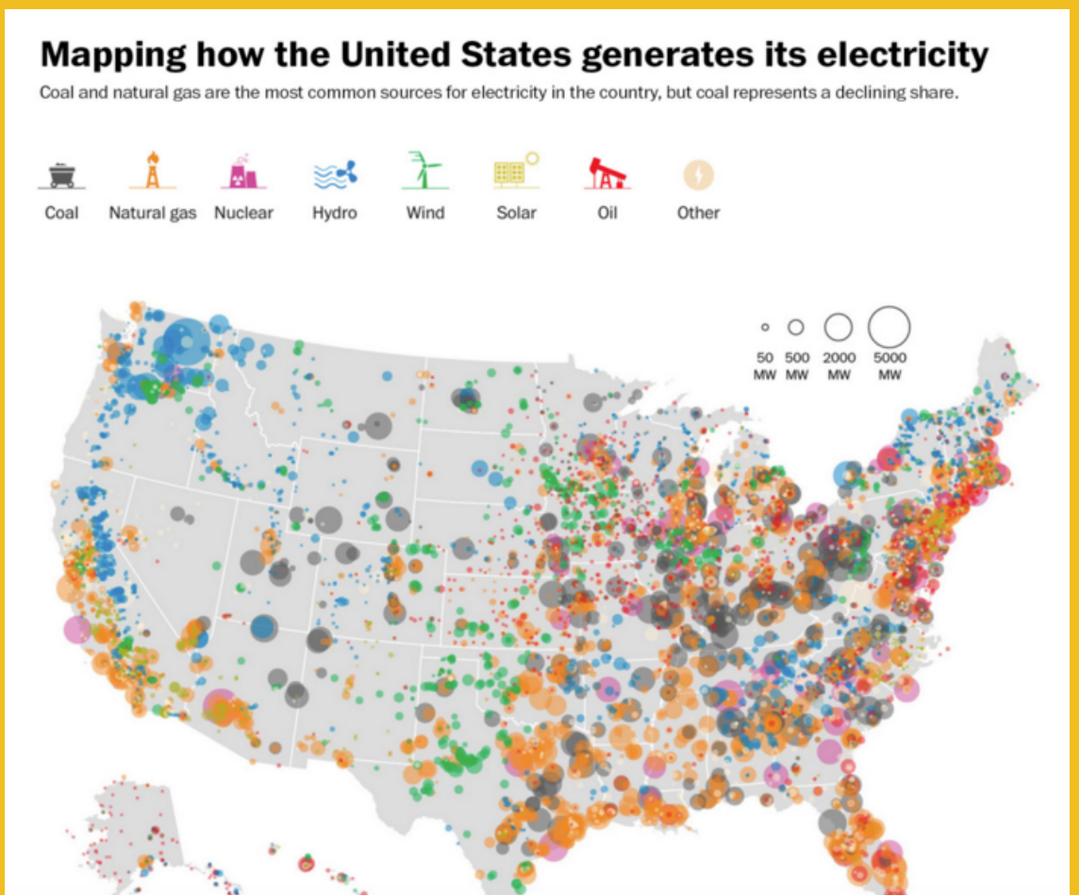
Explanatory Data Visualisation

> Focused on one or more key insights

> Greater element of storytelling in their production

> Less discovery-led

> Data Journalism tends to be more explanatory, for example



Exploratory ⟷ Explanatory

Washington Post
https://www.washingtonpost.com/graphics/national/power-plants/

Data Visualisation and Communication

Some Principles for Data Visualisation

> Choose the most appropriate and effective visualisations

> Always keep your audience in mind and do not deceive them simply to make a point

> A little design goes a long way

> Be careful with your choice of colour

> Choose your chart types with care

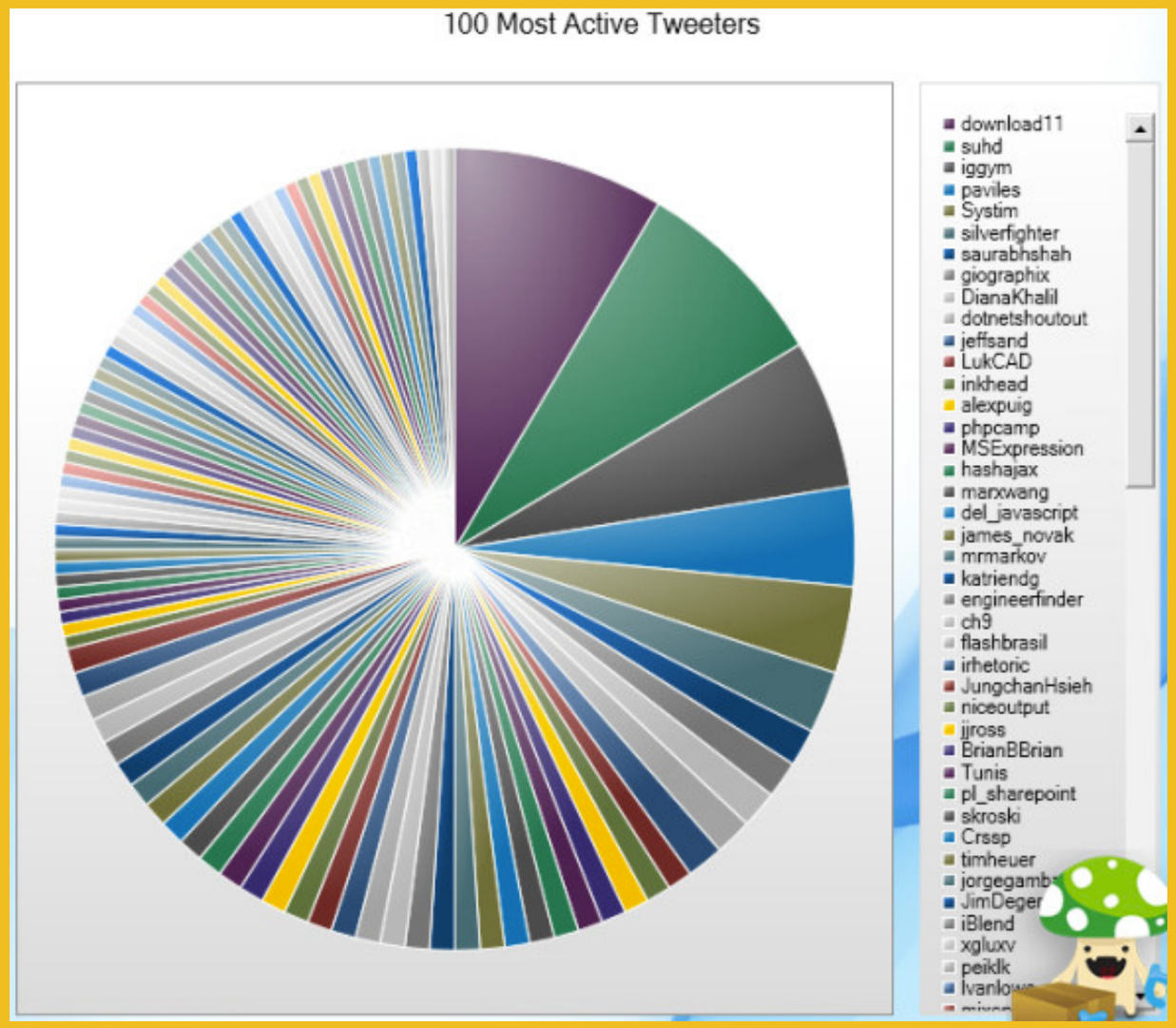
> In the vast majority of cases, do not mess with axes

> Don’t use 3D graphics and avoid pie charts

> Take care with proportional graphics

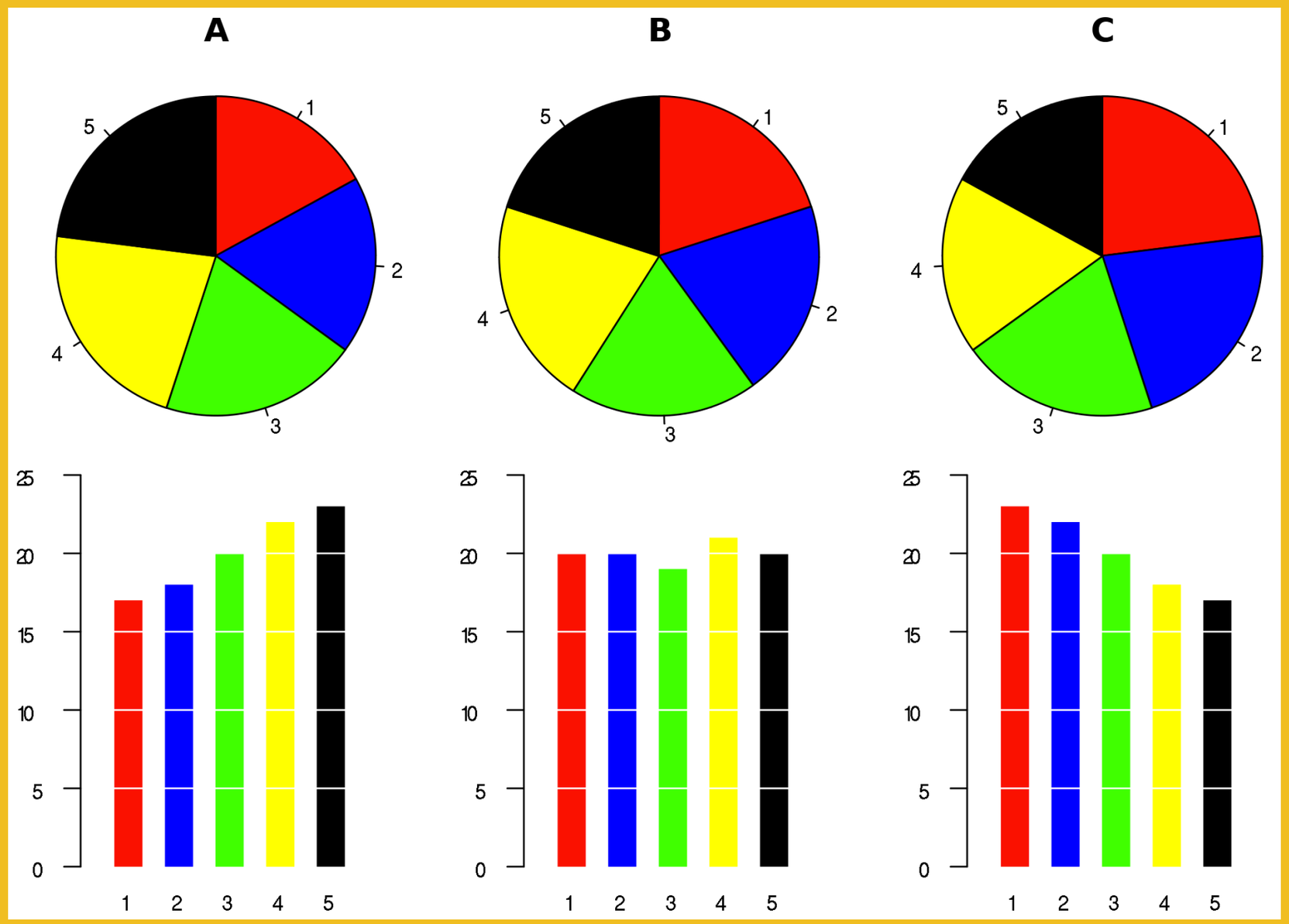
Data Visualisation and Communication

NO!



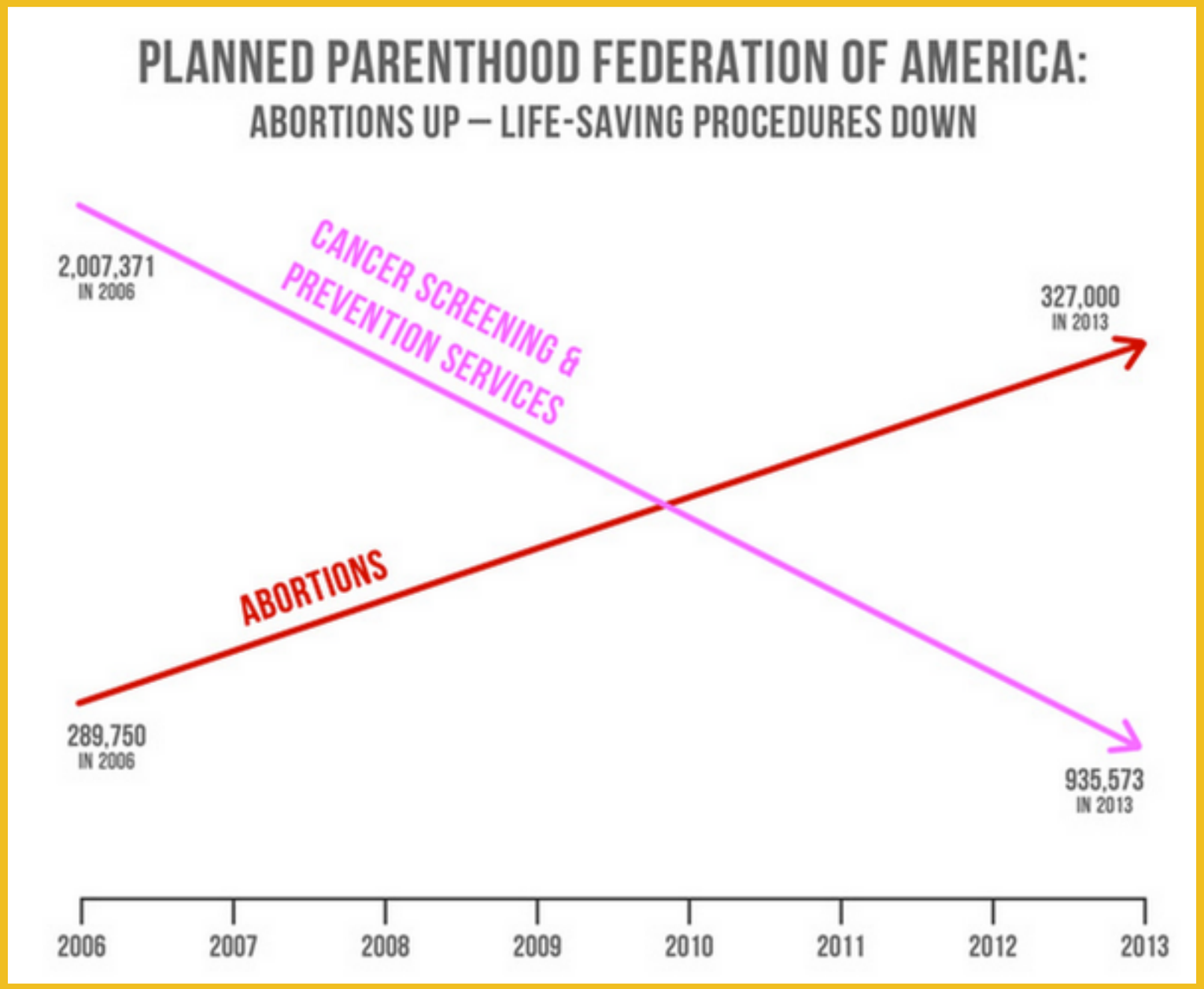
Data Visualisation and Communication

> Always consider other ways to communicate the story and evaluate your efforts



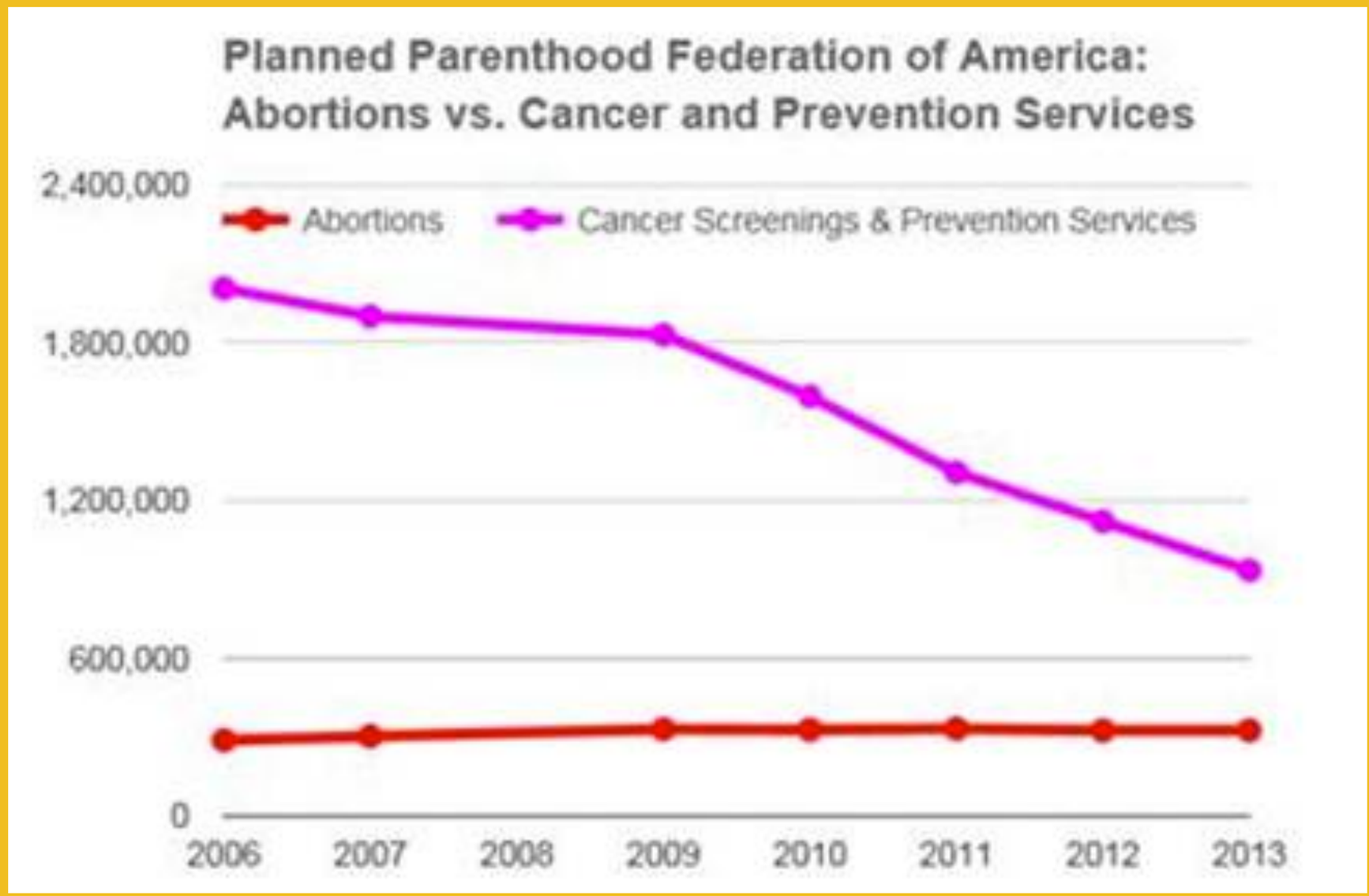
Data Visualisation and Communication

NO!



Data Visualisation and Communication

BETTER!



Data Visualisation and Communication

DEMO

Data Visualisation

We’re going to create a data visualisation dashboard that will help us understand Diabetes better on a local, national and international scale.

This demo uses Tableau, a data visualisation tool that is powerful and very intuitive. Feel free to follow along with the process or explore the code for yourself. Tableau Public is free and uni students can access more features by applying for a Tableau Desktop membership plan

End of Learning Lab, Part 2

✓ Exploratory Data Analysis

✓ Modelling

✓ Data Visualisation & Communication

+ableau®

Next…. INNOVATION LAB