

The Art of Data Visualization

<https://www.youtube.com/watch?v=AdSZJzb-aX8>

ADS ML – Week 3

Exploring Data with Python & Tableau

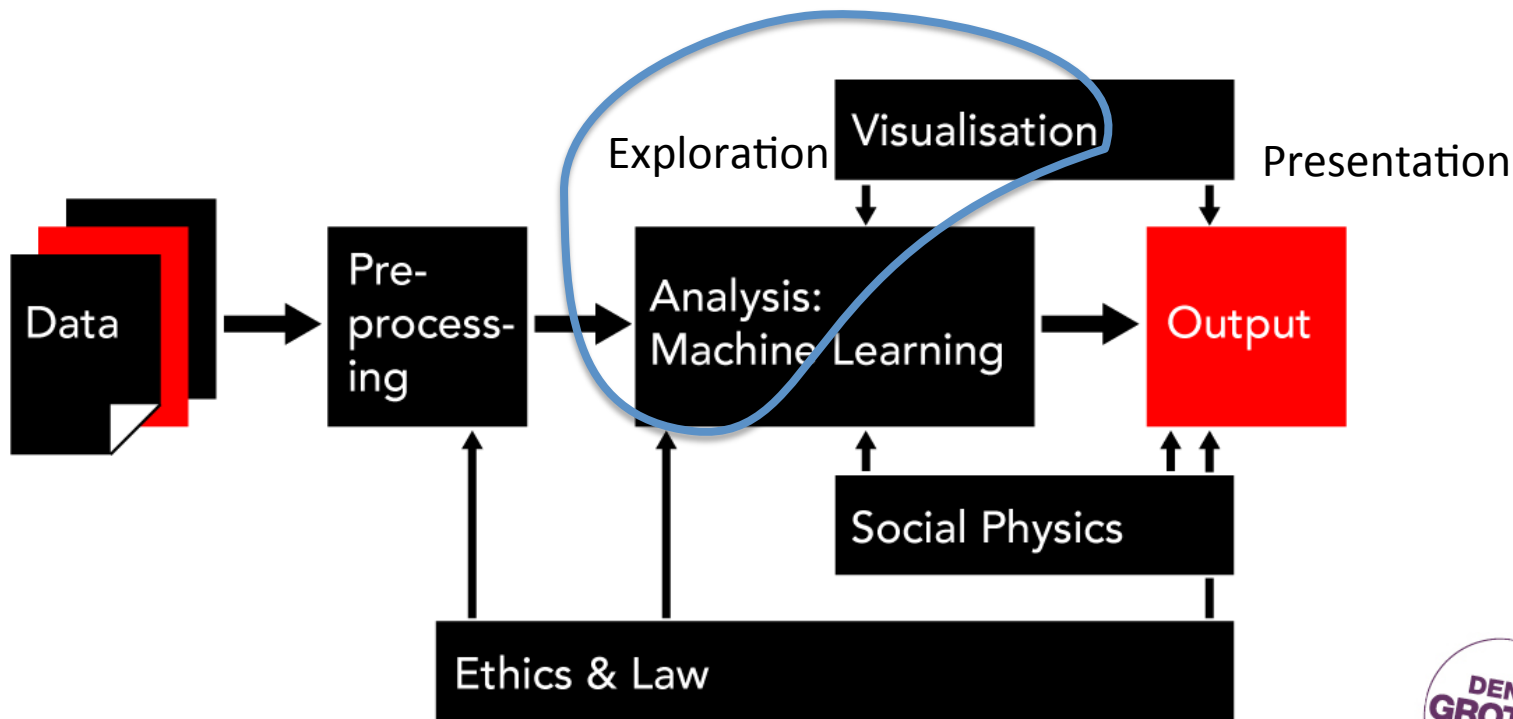
Goals: At the end of this lesson ...

- You can apply a few useful new exploration plot types.
- You learn to think about trends, outliers and clusters.
- You can use Tableau to explore data and answer questions in the form of a dashboard.

Contribution to the learning objectives

- **create** a visualization from any data set that is not misleading and that clearly shows clustering, outliers and trends,
- **motivate** every design choice in a created visualization,
- **motivate** the next step in a data analysis based on a given visualization,
- **present** data analysis and visualizations as part of reproducible research,
- **apply** narrative techniques in visualizations,
- **create** engaging visualizations that allow for data exploration and story progression.

What are we doing?



Outliers, trends and clusters

Outliers, trends and clusters.

Data exploration usually starts by asking questions about your data.



Outliers

- What is the best/highest ...?

Trends

- How does ... change with ...?

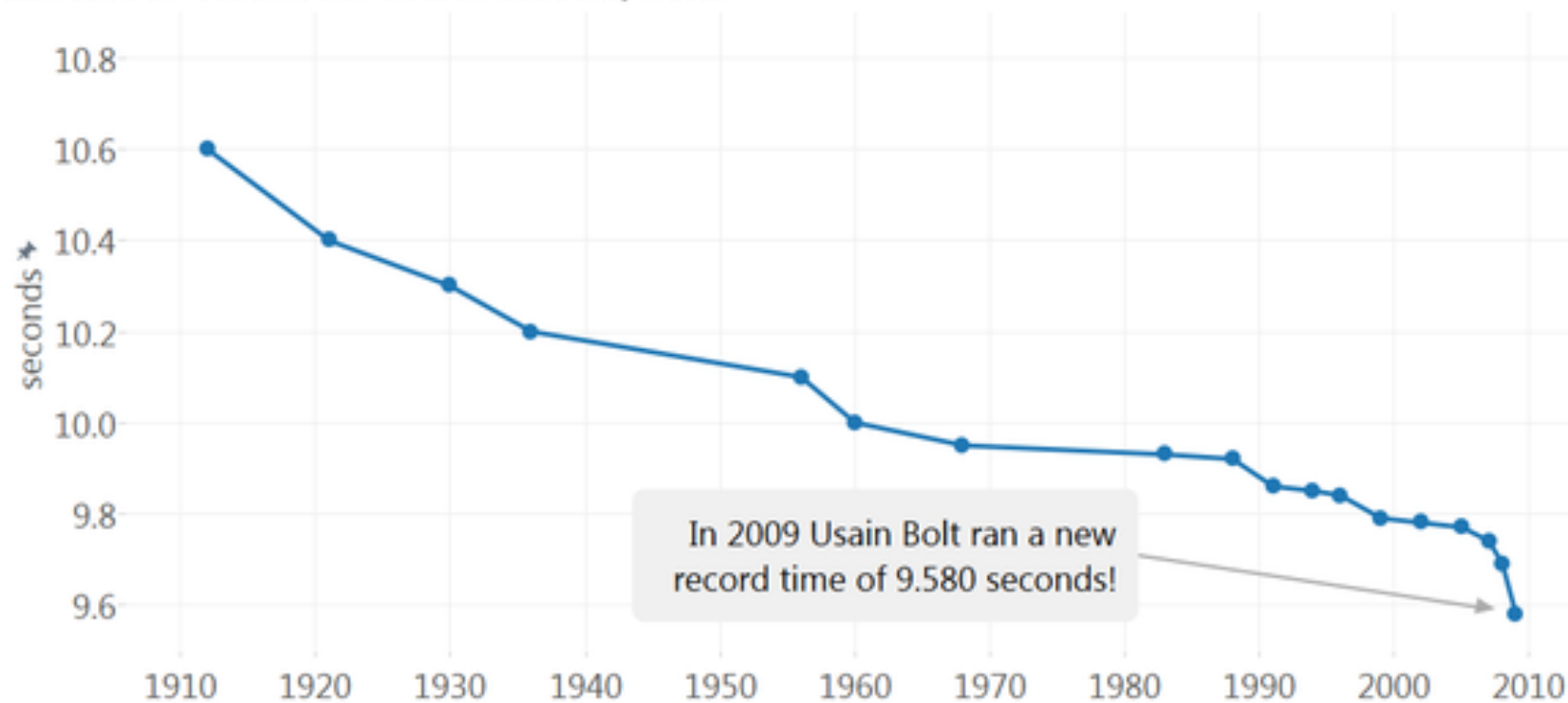
Clustering

- How is ... distributed?

How has the 100m world record changed since 1910?



Usain Bolt is an #outlier: he blew the 100m record away in 2009.



Gotta Viz Em All



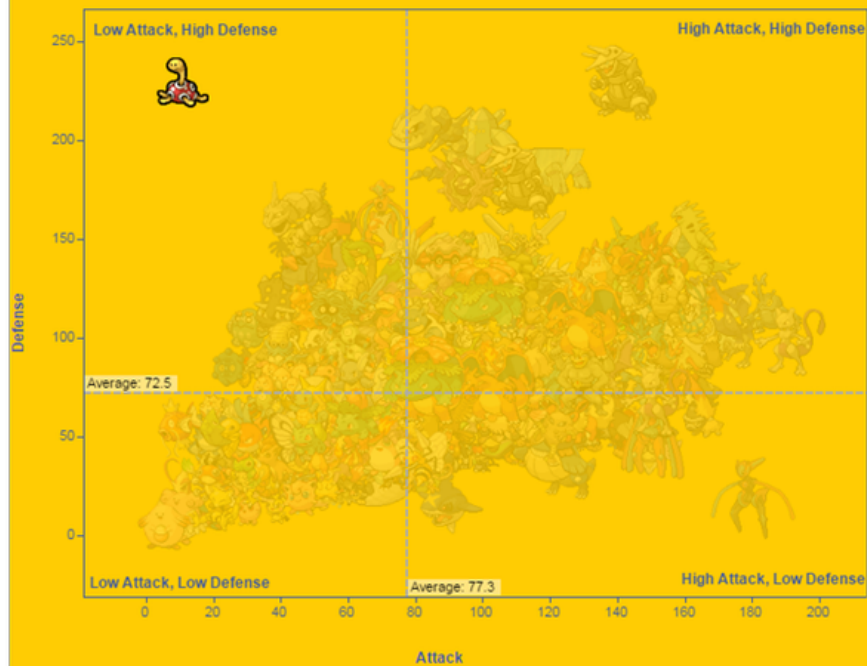
Choose Metric for X-Axis

Choose Metric for Y-Axis

Find a Pokemon

Attack

Defense



Rivers of the World: Scale Rank 1-9

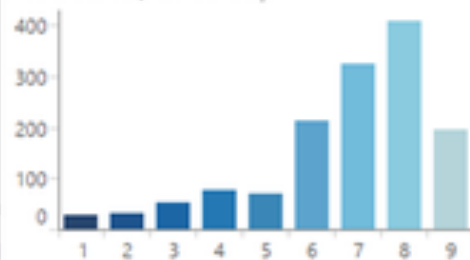
Find River by Name

Scale Rank



River Scale Rank

click a bar to filter the map



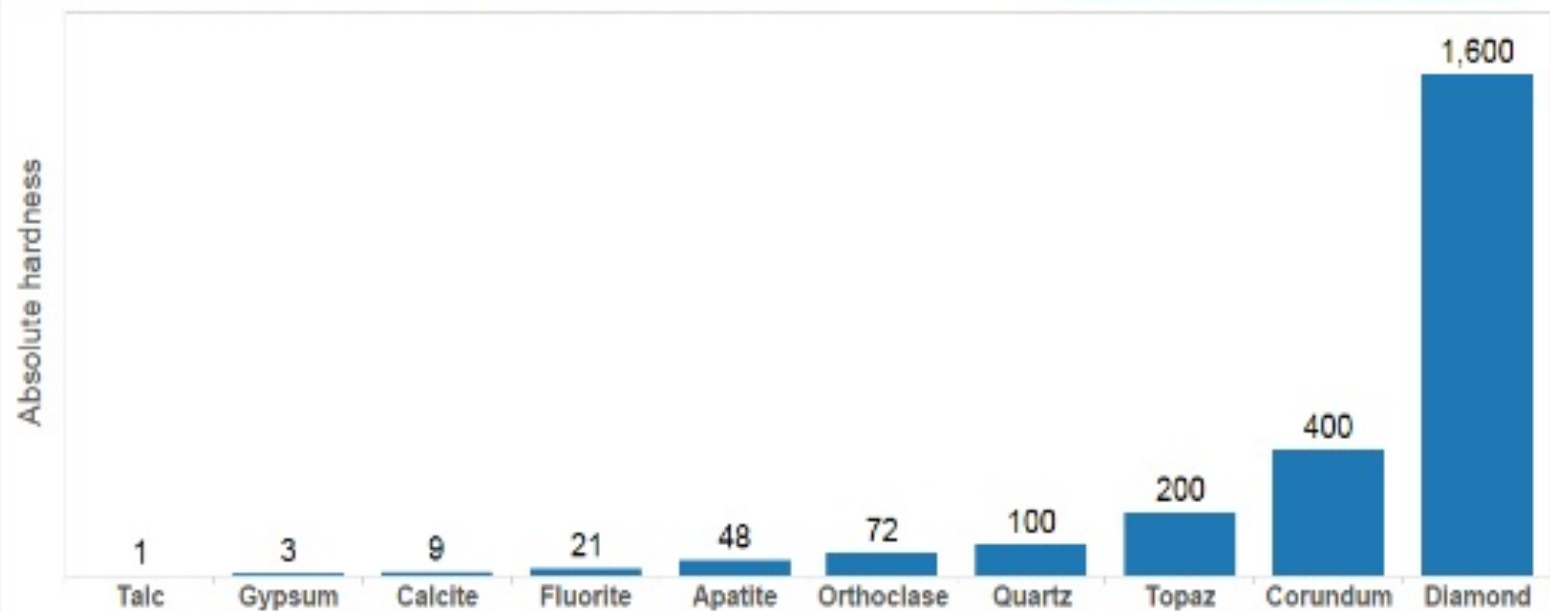
"Scale Rank" indicates the zoom level at which each river should start being shown. Scale Rank 1 is for major rivers like the Nile.

Scale Rank



Diamond is an #outlier with an absolute hardness of 1600

Data: http://en.wikipedia.org/wiki/Mohs_scale_of_mineral_hardness



Data exploration in Python

Data exploration in Python

- We have seen the basic plots like scatter, bars and histograms.

Find out by doing assignment 1 of this week. (30min)

<https://github.com/olafjanssen/ads-dv>

- Do you know the useful plotting charts: **parallel coordinates** and **scatter matrix**?

Tableau

Plotting with Tableau

- Tableau is a very popular visualization tool among BI people and data analysts.
- It is very powerful for visualization with a design-rule friendly workflow.
- Target audience is usually not the general public.

- <http://www.tableau.com/>





Why Tableau?

- Allows for connecting with Big Data sources such as Spark and MongoDB.
- Simple for the exploration phase.
- Allows also for publishing online interactive visualizations with storytelling elements.



Fonttys Analytics interface showing a data source named "Voting Final (eurovision...)" and a visualization setup.

Data | **Analytics**

Dimensions

- Country
- Giver
- Year
- Measure Names

Measures

- Score
- Latitude (generated)
- Longitude (generated)
- Number of Records
- Measure Values

Pages

Columns

Rows

Filters

Marks

Automatic

Color, Size, Text, Detail, Tooltip

Drop field here

Drop field here

Drop field here

Drop field here

Data Source | Sheet 1

Assignments 2 - 4

- Practice with Tableau using predefined tutorials.
- Create a dashboard using your own questions and dataset of choice.
- (Optional) Recreate it in iPython as part of reproducible research.

Before week 4:

Watch lectures **1a** en **2a** of the Udacity MOOC **Data Visualization and D3.js** and create your own cheat sheet, with all the design rules and tips!

Helpful extra info:

<http://www.slideshare.net/Visage/data-visualization-101-how-to-design-chartsandgraphs>