Data Mining

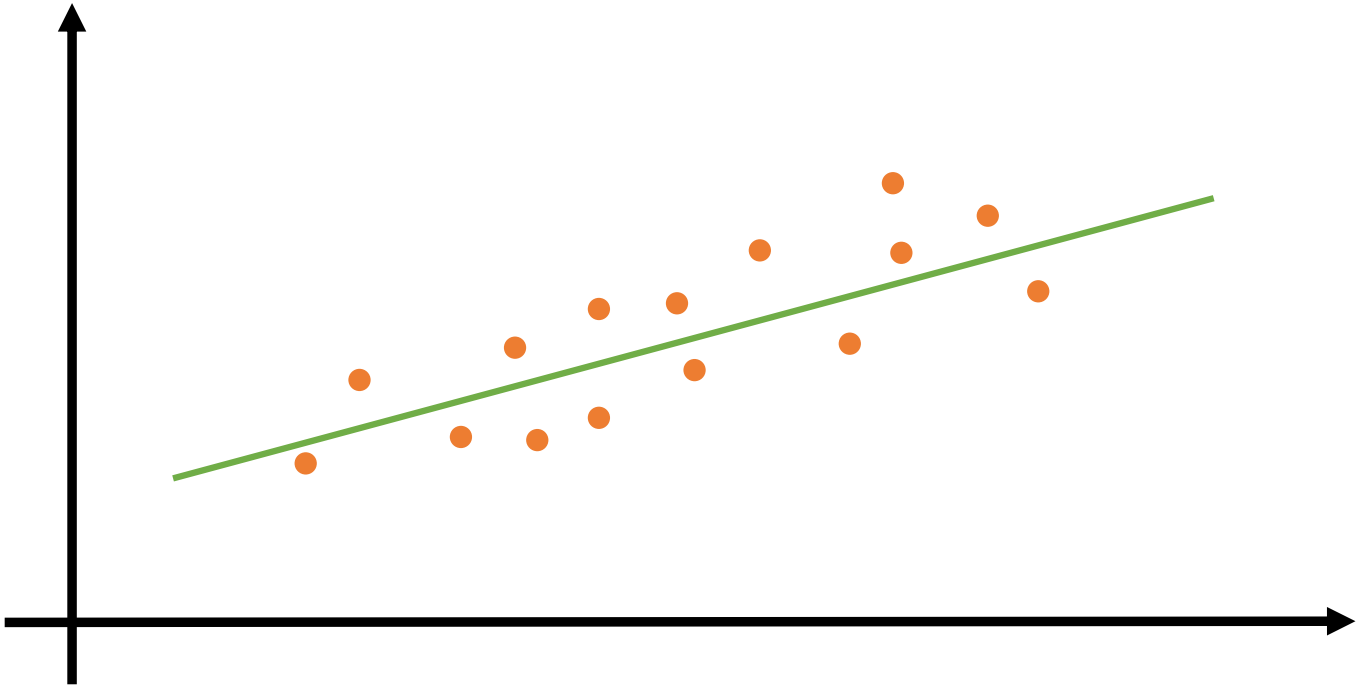
COMP23111 – Database Systems

Gareth Henshall
Lecturer in Computer Science

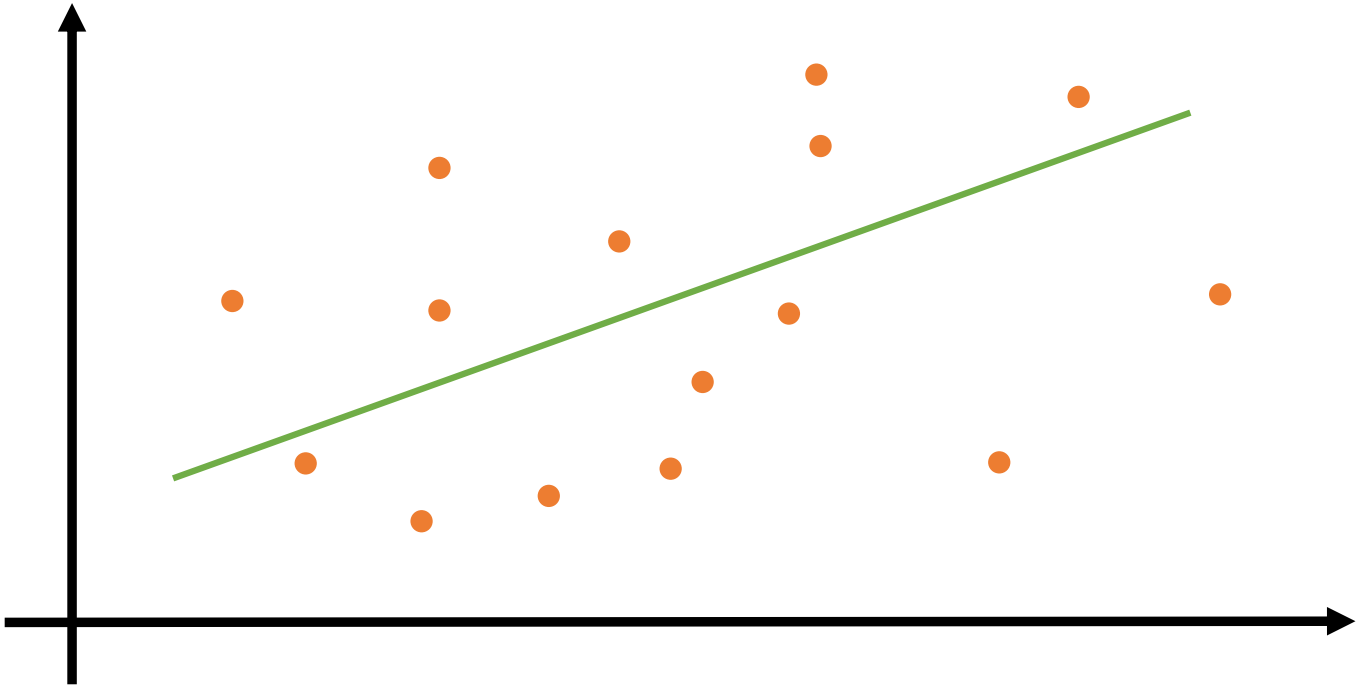
What is Data Mining?

The process of discovering trends and patterns in large sets of data

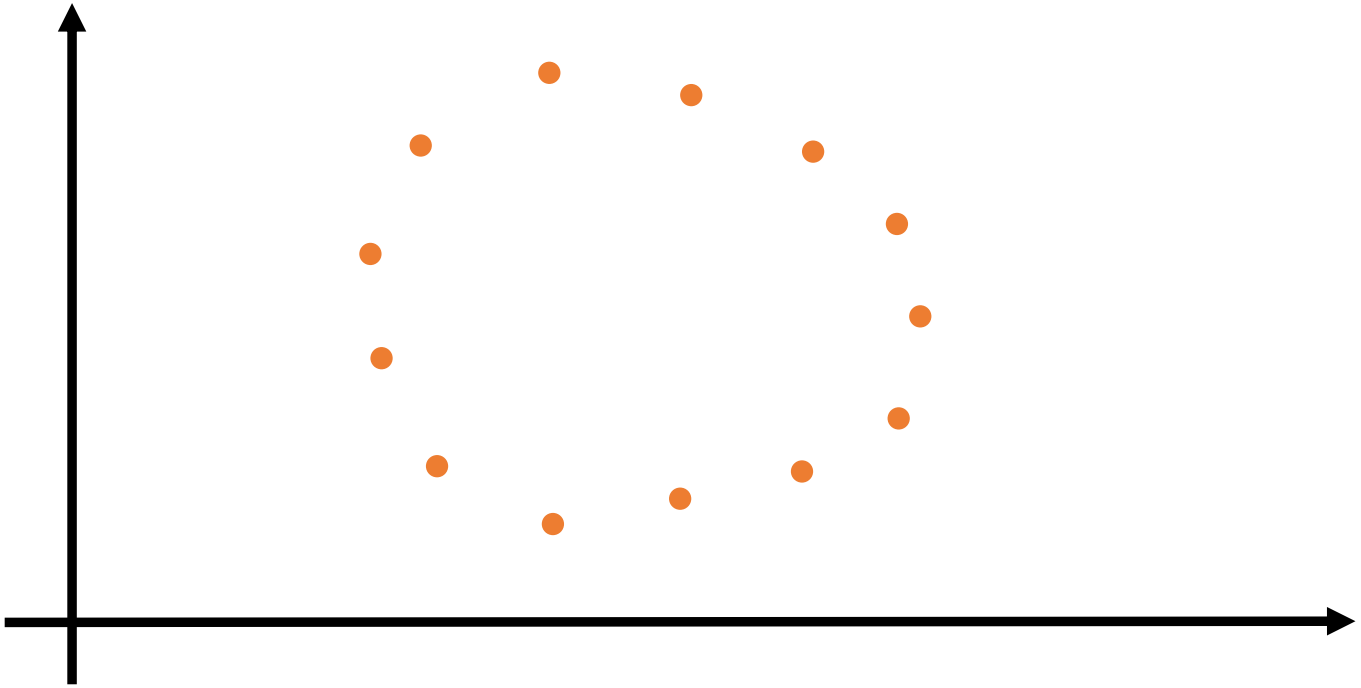
Spotting the Patter



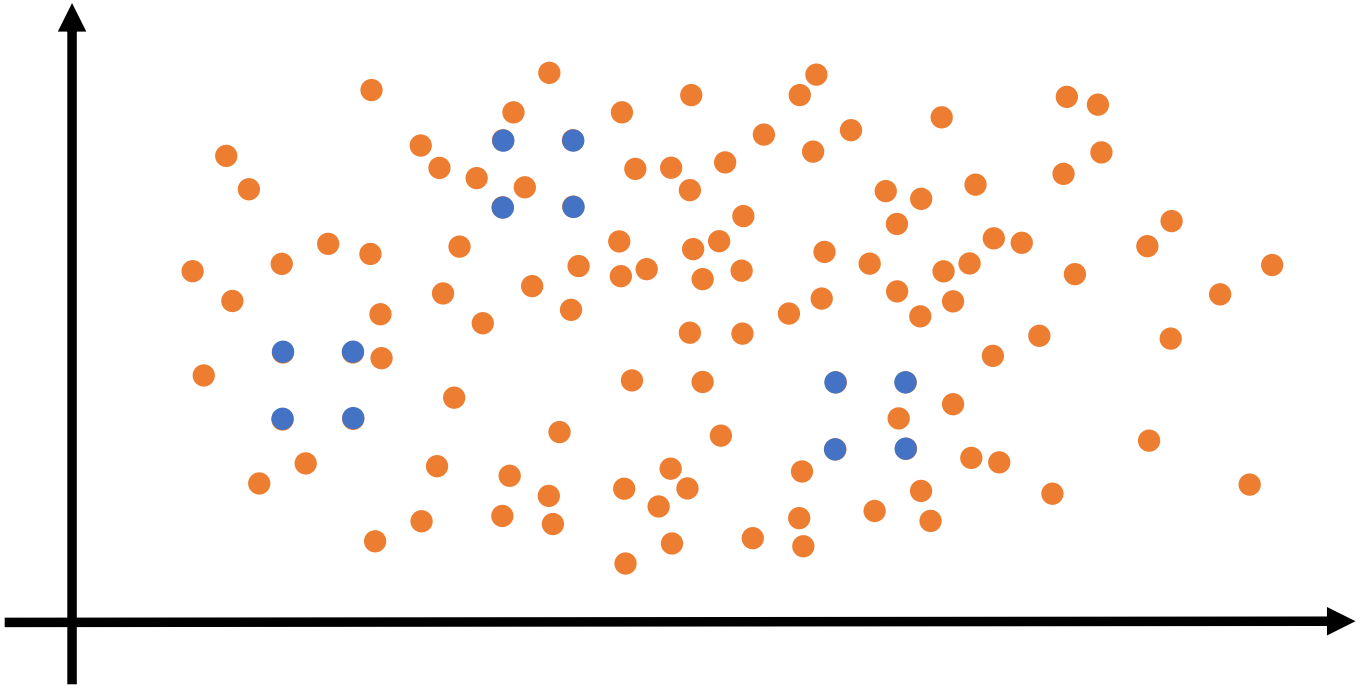
Spotting the Patter



Spotting the Patter



Spotting the Patter



Data Pre-Processing

Real world data can be:

Incomplete

Missing attribute values:
`occupation = " "`

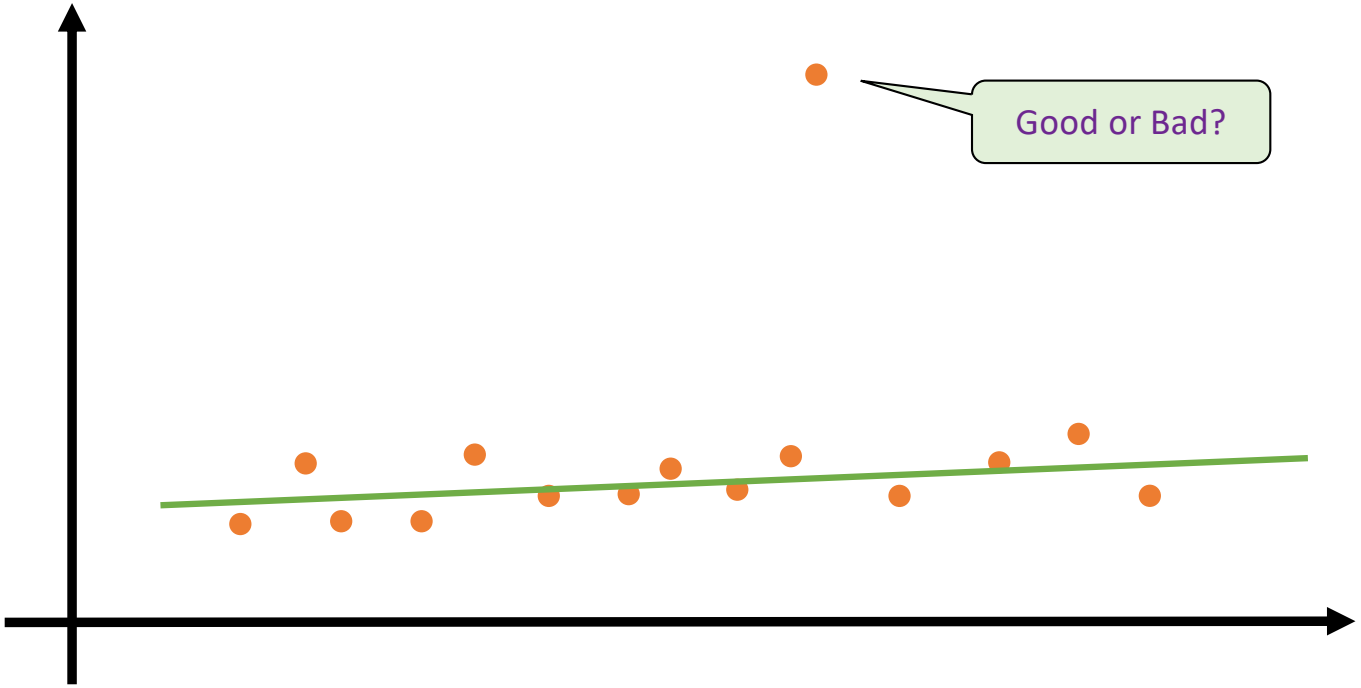
Noisy

Contains errors:
`salary = "-1"`

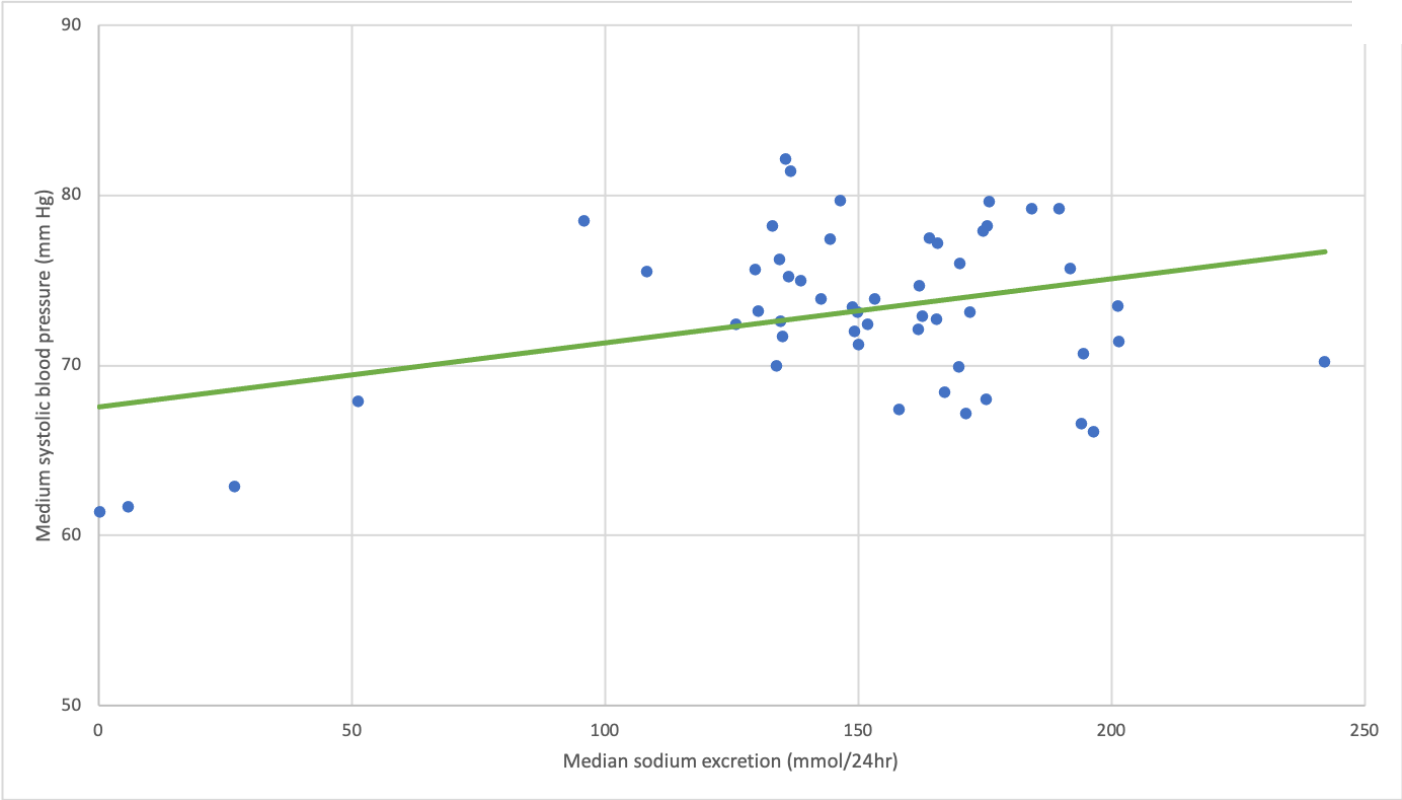
Inconsistent

Contains discrepancies:
`age = "20"`
`dob = "01/01/1990"`

Eliminating Background Noise



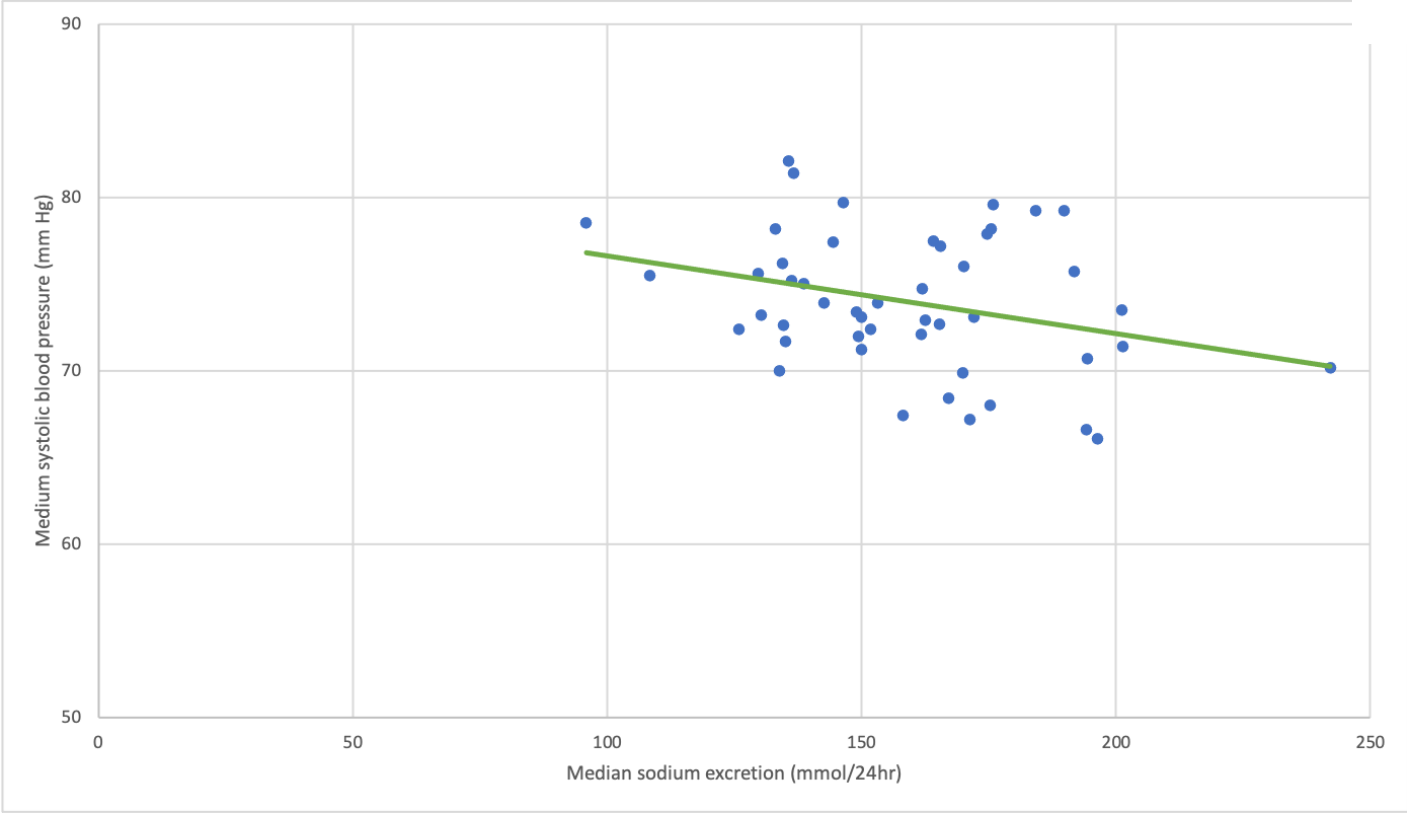
Reading Data Wrong?



Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. 1988

https://www.jstor.org/stable/29700360?seq=1#metadata_info_tab_contents

Reading Data Wrong?



Intersalt: an international study of electrolyte excretion and blood pressure. Results for 24 hour urinary sodium and potassium excretion. 1988

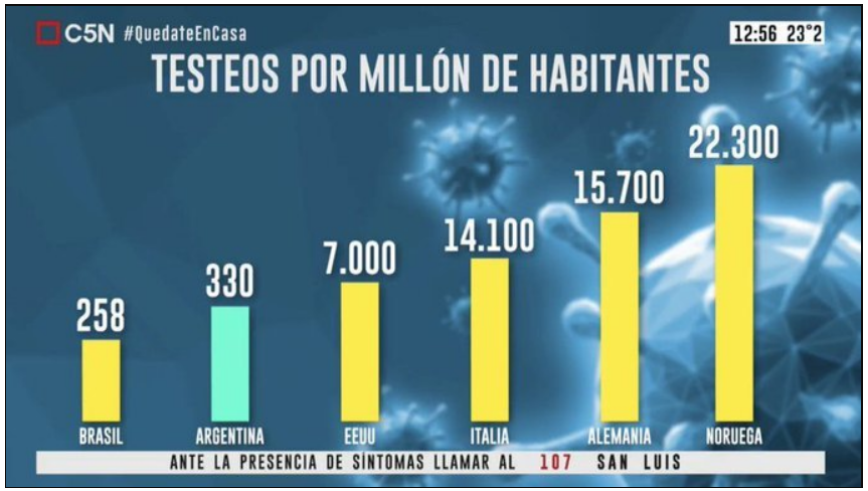
https://www.jstor.org/stable/29700360?seq=1#metadata_info_tab_contents

DISCLAIMER!

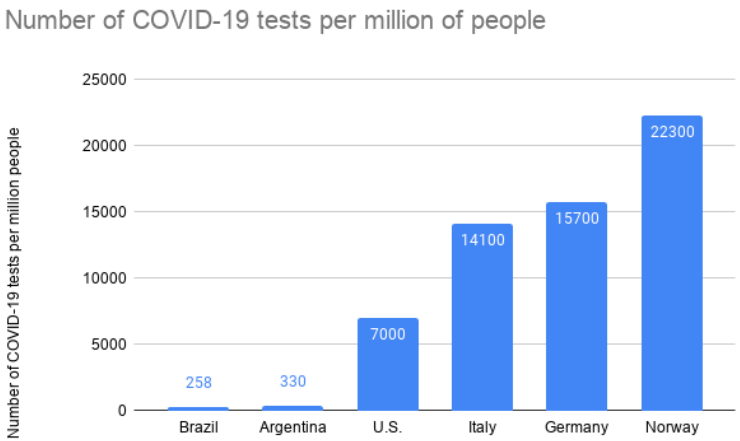
Over the next few minutes we will be discussing COVID-19, if you have been affected by COVID-19 and think you may find the material upsetting stop the vid now.

Data in the Media

Argentinian TV channel C5N



Same graph but with Y Axis

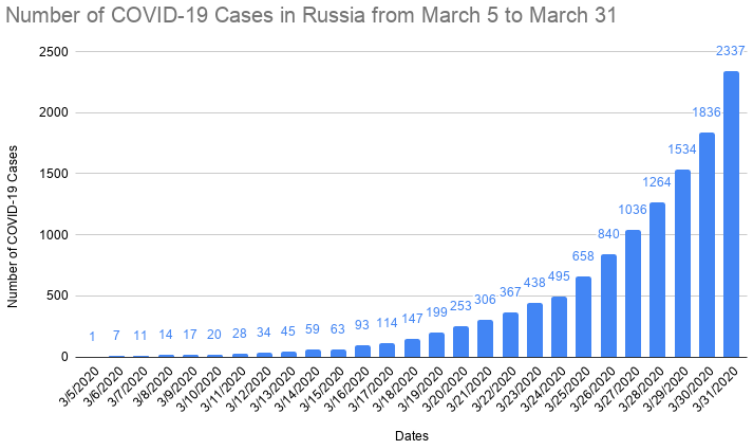


https://towardsdatascience.com/stopping-covid-19-with-misleading-graphs-6812a61a57c9

Data in the Media

Russia Flattening the Curve

The Actual Curve



Data in the Media

Georgia Department of Public Health