

Introduction to Data Science

Hello World!

Joanna Bieri DATA101

Intro

Important Information

- Email: joanna_bieri@redlands.edu
- Office Hours: Duke 209 [Click Here for Joanna's Schedule](#)

Announcements

- **Please come to office hours to get help!**
- Be patient and kind with yourself and others. Sometimes getting started can be overwhelming, **DON'T GIVE UP**, you can do this and I can help!
- I can be **REALLY FLEXIBLE** with deadlines in the first few weeks of class, so if you run into technology issues please don't panic.

Day 2 Assignment

OLD 1. Make sure you can **Fork** and **Clone** the Day2 repo from Redlands-DATA101

NEW 1. Make sure **Pull** any new content from the class repo - then **Copy** it over into your working directory.

- 3 Open the file Day2-HW.ipynb and start doing the problems.
 - You can do these problems as you follow along with the lecture notes and video.
- 4 Get as far as you can before class.
- 5 Submit what you have so far **Commit** and **Push** to Git.
- 6 Take the daily check in quiz on **Canvas**.
- 7 Come to class with lots of questions!

Check out the JupyterLab Notebook for Today!

- Day2-HW.ipynb contains follows along with the lecture notes and video!

Data Science

Let's do some Data Science!

This exploration is from <https://datasciencebox.org/> and the original author of this content (written in R) is Mine Çetinkaya-Rundel.

I have updated it for our class and translated the code to python.

Let's do some Data Science!

Introduction

How do various countries vote in the United Nations General Assembly, how have their voting patterns evolved throughout time, and how similarly or differently do they view certain issues?

Let's do some Data Science!

Data

The data we're using originally come from the `unvotes` R. package. This package provides the voting history of countries in the United Nations General Assembly.

Exploratory Data Analysis

- Your first job with any data set is to try to understand what is actually there.
 - How much data do you have?
 - What does the data means?
 - How was it collected?
 - Are there any privacy concerns?
 - Maybe take some averages or other basic statistics.

Visualization

Data Visualization

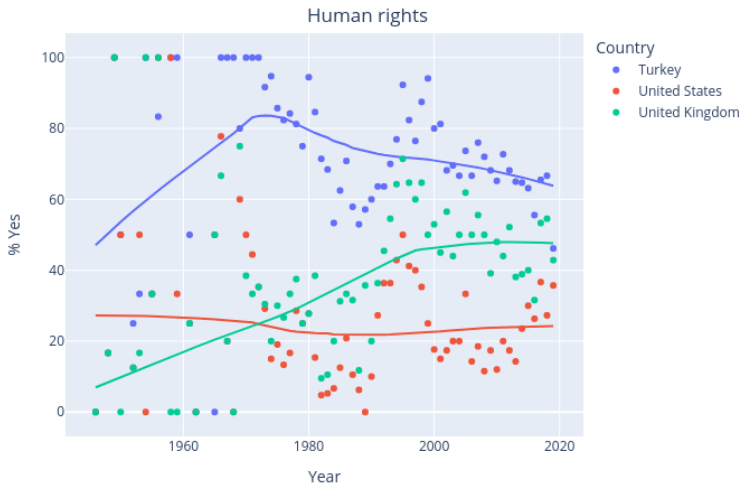
Visualization is a great way to get to know your data and tell a story!

Let's create a data visualization that displays how the voting record of the US changed over time on a variety of issues, and compares it to two other countries: UK and Turkey.

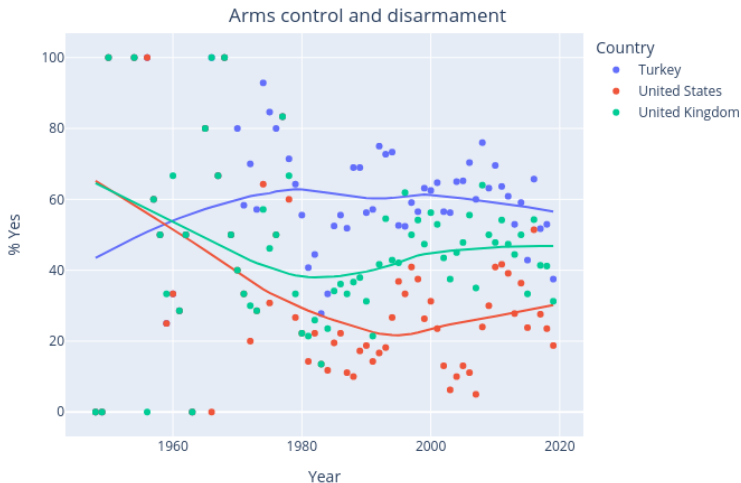
Data Visualization

- Scatter plot - Dots represent each data point
- Trend line - LOWESS (Locally Weighted Scatterplot Smoothing) - creates a smooth line through a scatter plot to help you to see relationship between variables and foresee trends. A type of regression
- Always label your axes!

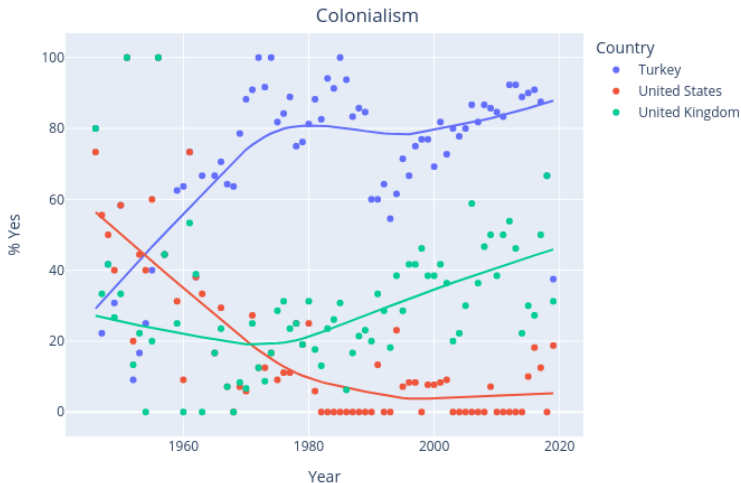
Data Visualization - Human Rights



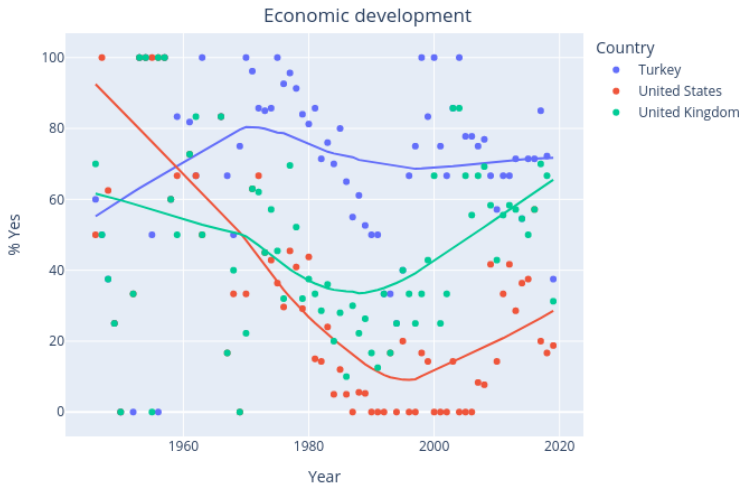
Data Visualization - Arms Control



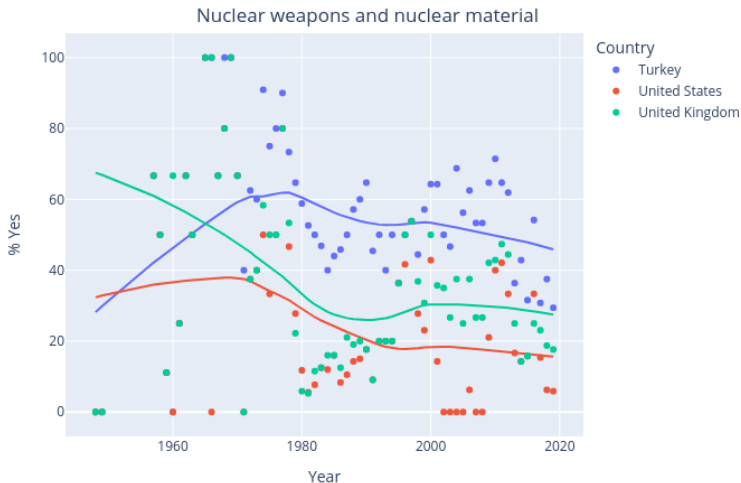
Data Visualization - Colonialism



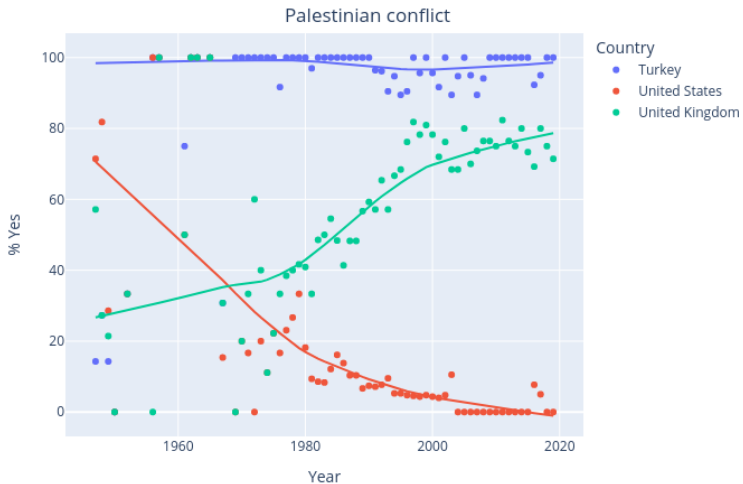
Data Visualization - Economic Development



Data Visualization - Nuclear Weapons



Data Visualization - Palestinian Conflict



Explore the Data!

Now it is your turn to explore the data!

Questions?