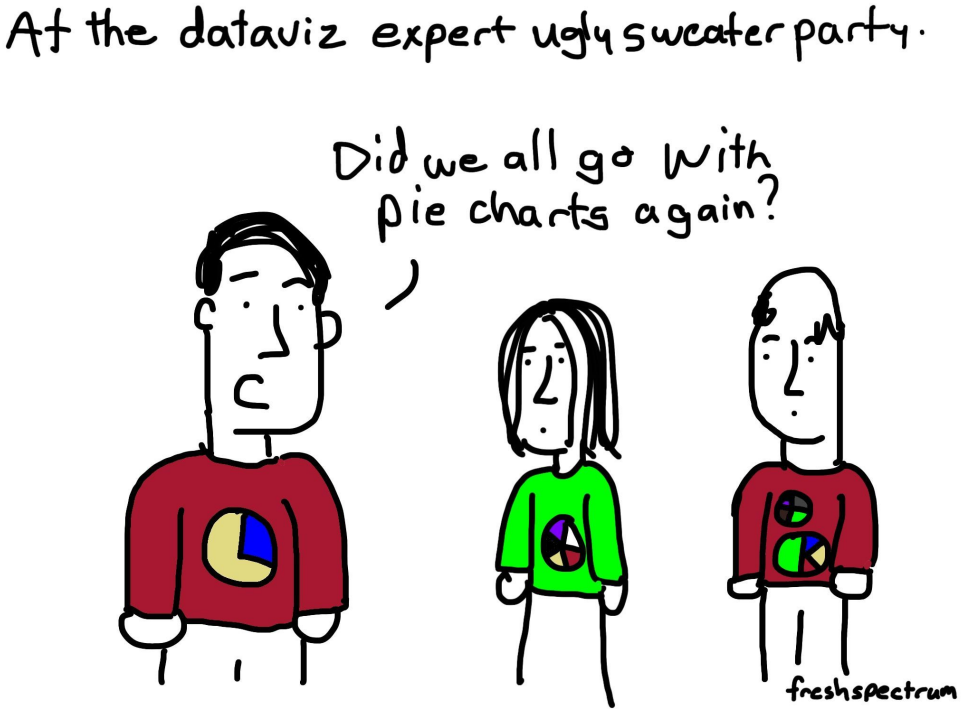

At the dataviz expert ugly sweater party.
Did we all go with pie charts again?
freshspectrum

Source

Welcome to Week 4 Lecture 1!

Data Science in Python & Machine Learning

Learning Goals:

By the end of this less you will be able to:

1. Know the end-of-stack expectations

2. Explain the difference between exploratory and explanatory visualizations

3. List 3 considerations when making an explanatory visualization

4. Create and interpret a visualization and answer a specific question.

Week 4 CORE Assignments

1) Project 1 - Part 4 (Core)

Must be submitted by <u>Friday, February 18th @ 9 AM PST</u>

- ***Any requested Project 1 - Part 4 resubmissions will NOT prevent you from moving on to Stack 2 next week!***

In addition to Passing the Belt Exam, ALL Core Assignments must be completed by Friday 02/18 by 9 AM PST!

Source

📣Stack 2 Announcement

- Coding Dojo's switching from a single instructor teaching all stacks to each instructor specializing in specific stacks.

- For Stack 2, I will be handing over the reigns to Sherlin Whaley, our Machine Learning Instructor!

- **I will continue to offer bonus Friday office hours/lectures on Fridays during Stack 2.**

Explanatory Visualizations Tell a Story

Image Source

Presentation Graphs Should:

- Have clear answer to a well defined question
    - ‘Is there a correlation between heat and humidity?’
    - ‘Are men or women more likely to buy our product?’
    - ‘How have searches for Pokemon changed over the last 10 years?’
- Be clear enough to understand in 5 - 10 seconds
- Be visually appealing, but NOT cluttered or distracting
- Use shape, color, words, and/or markers to draw attention to what is important

GDP Change Over 50 Years

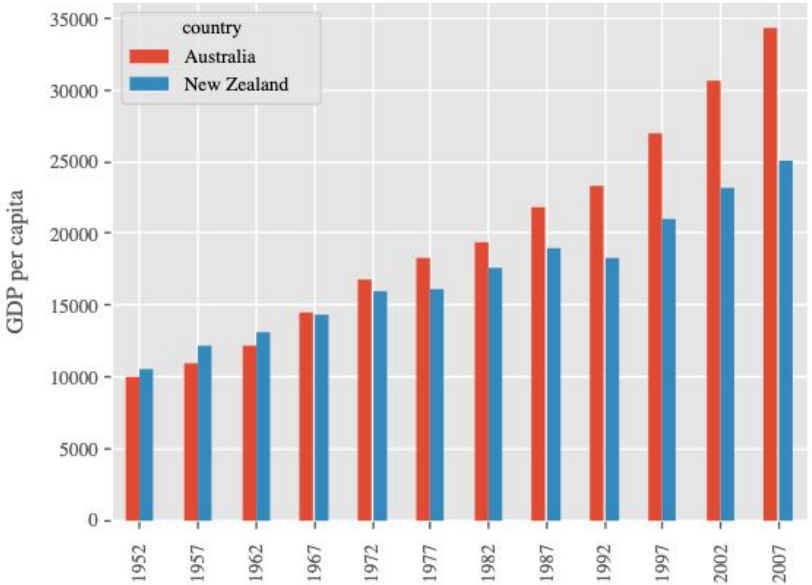


Image courtesy of Software-Carpentry

Know your Audience

Consider:

- Technical understanding
- Necessary amount of detail
- Content of interest

Know your Data

Consider:

- Meaning of your features (columns)
- Relationships between features
- Types of data (numeric or categorical)

Know your Data

Data Relationship	Chart Type
Correlation	Scatter Plot
Distribution	Histograms, Box Plots
Geospatial	Maps
Ranking or Comparison	Bar Charts
Trends	Line Graphs

Know your Goal

Consider:

- What question does the image answer?
- What you want your audience to get out of your visual?

Know your Resources

- Selecting Visualizations
  - https://www.python-graph-gallery.com/
  - https://datavizproject.com/

- Knowing What is Possible - Example Galleries
  - https://seaborn.pydata.org/examples/index.html
  - https://matplotlib.org/stable/gallery/index.html

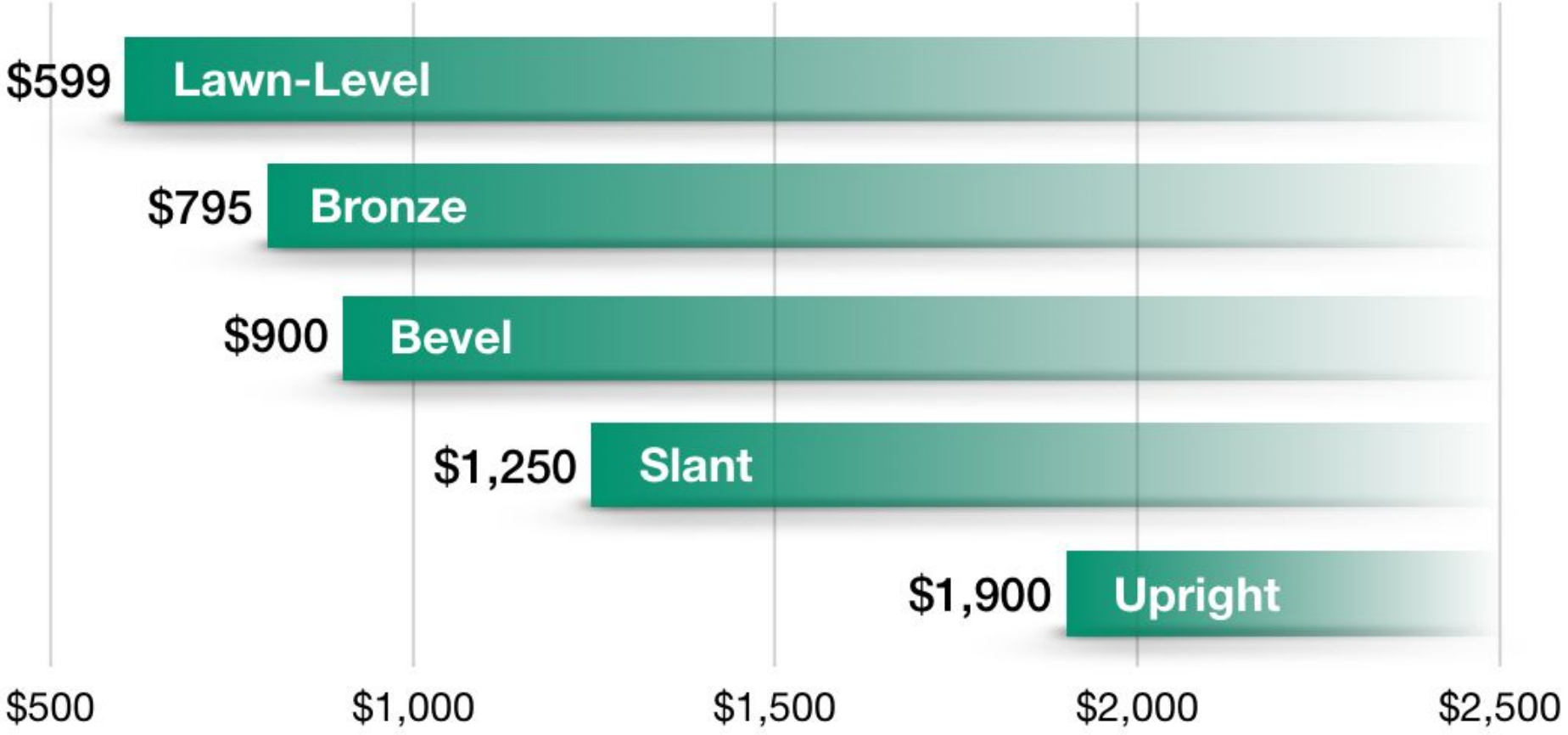
Explanatory Visualizations Tell a Story

Spend 5 minutes or so to discuss how to improve these graphs.


Map win statistics
100
80
60
40
20
0
Win %
15
12
9
6
3
0
Times Played
Inferno
Mirage
Overpass
Train
Nuke
Dust2
Win %
CT Win %
T Win %
Times played
All stats for Natus Vincere

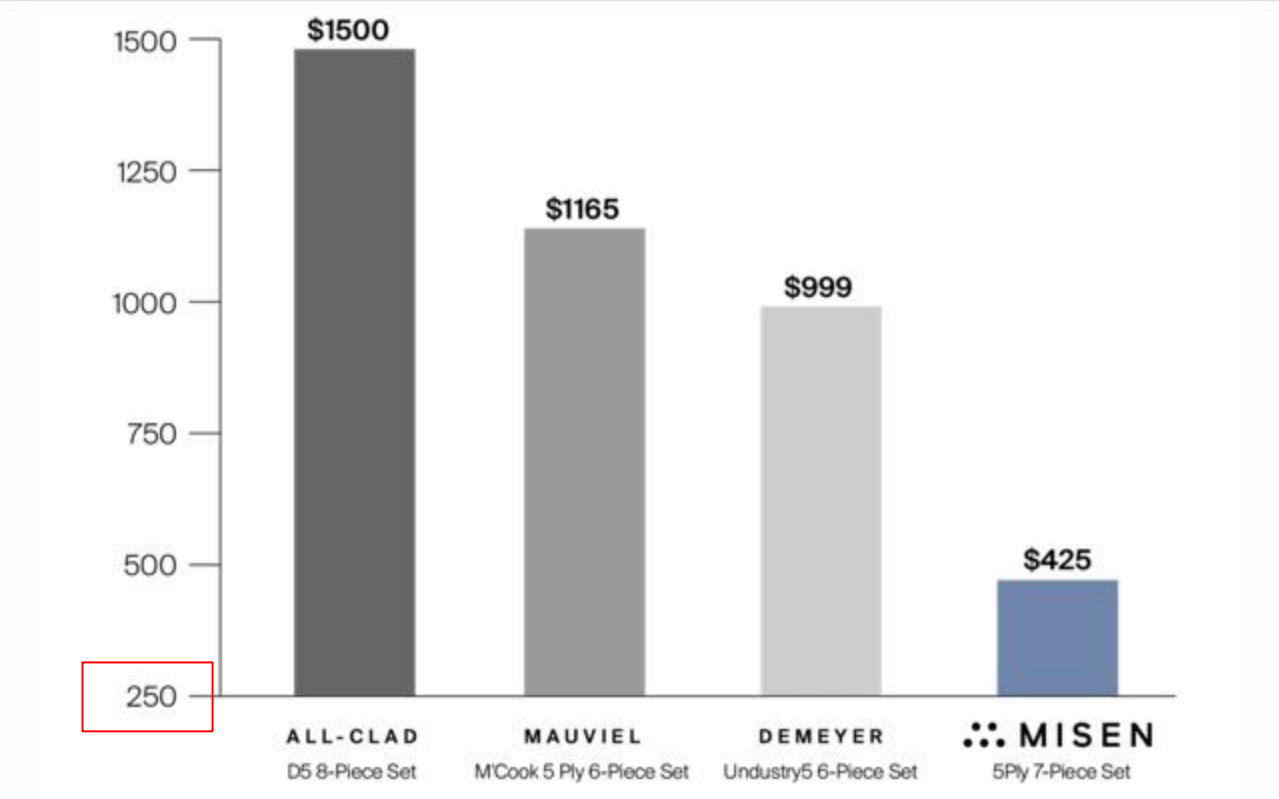
Less is More

- Keep your visualization as simple as possible
- Understandable in 5-10 seconds


$599
Lawn-Level
$795
Bronze
$900
Bevel
$1,250
Slant
$1,900
Upright
$500
$1,000
$1,500
$2,000
$2,500

Do not mislead your audience

- Always go from left to right

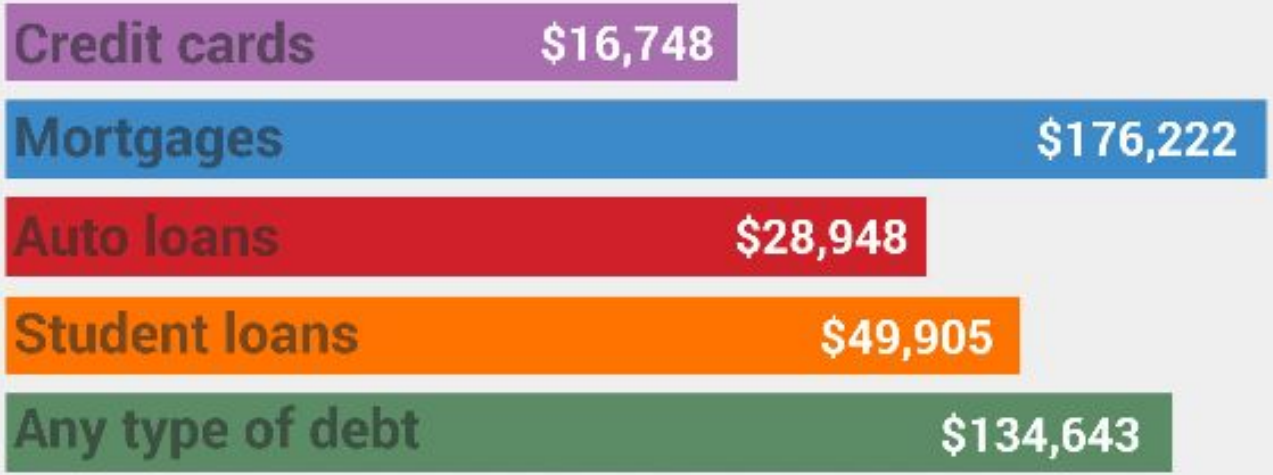

1500
1250
1000
750
500
250
$1500
$1165
$999
$425
ALL-CLAD
D5 8-Piece Set
MAUVIEL
M'Cook 5 Ply 6-Piece Set
DEMEYER
Undustry5 6-Piece Set
MISEN
5Ply 7-Piece Set

Do not mislead your audience

- Do not break axis, always start at 0

Types of debt

The total owed by the average U.S. household, by debt type.



Do not mislead your audience

- Keep sizes proportional to their values

HOW MUCH DO YOU SPEND ON GROCERIES EVERY WEEK?

22% UNDER $100

26% ABOUT $100

39% $100 TO $200

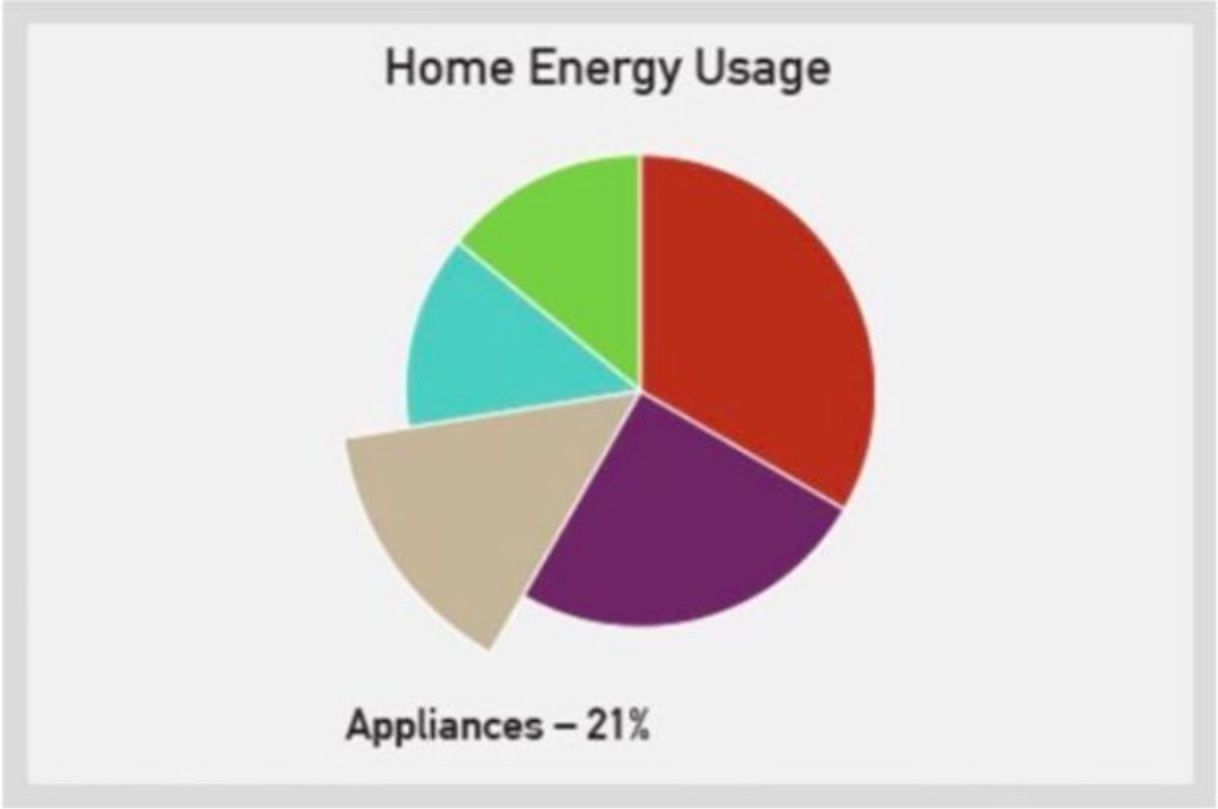
10% $200 TO $300

3% MORE THAN $300

@THEKITCHN

Do not use data visualization to express your creativity!

- Keep it simple and clear!


Home Energy Usage
Appliances – 21%

Do not mislead your audience

- Stay away from making the reader estimate volumes (avoid 3D graphs)
- Stay away from shapes that are hard to estimate areas (yes, this means avoiding pie charts)

Schooling and Life Expectancy - CodeAlong

Let's say I want to report on the relationship between education and life expectancy…

The Data:

- This dataset shows various statistics for hundreds of countries for 15 years. Each row is one country during one year.

- The data is available on Kaggle.

Image source

Schooling and Life Expectancy CodeAlong: Questions to Answer

1. **Does education increase life expectancy?**
2. **Does the development status of a country affect schooling or life expectancy?**
3. **How have schooling and life expectancy changed over time?**

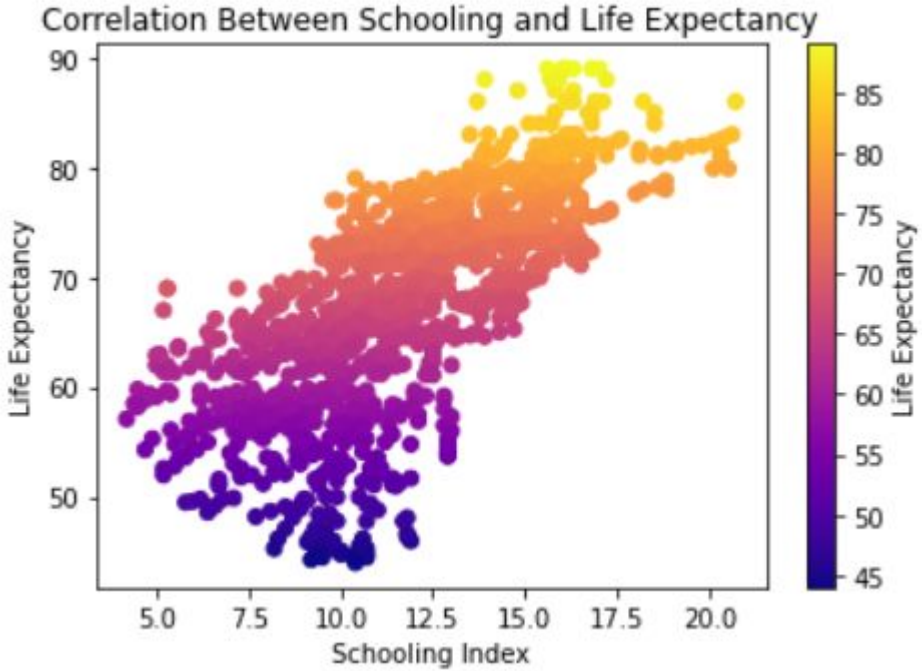
[Our Starter Colab Notebook]

Schooling and Life Expectancy

Does education increase life expectancy?

Schooling and Life Expectancy

Does education increase life expectancy?



Schooling and Life Expectancy

Does education increase life expectancy?

Interpretation:

This scatter plot shows a positive correlation between schooling and life expectancy. Nations with better schooling tend to have longer life expectancy.

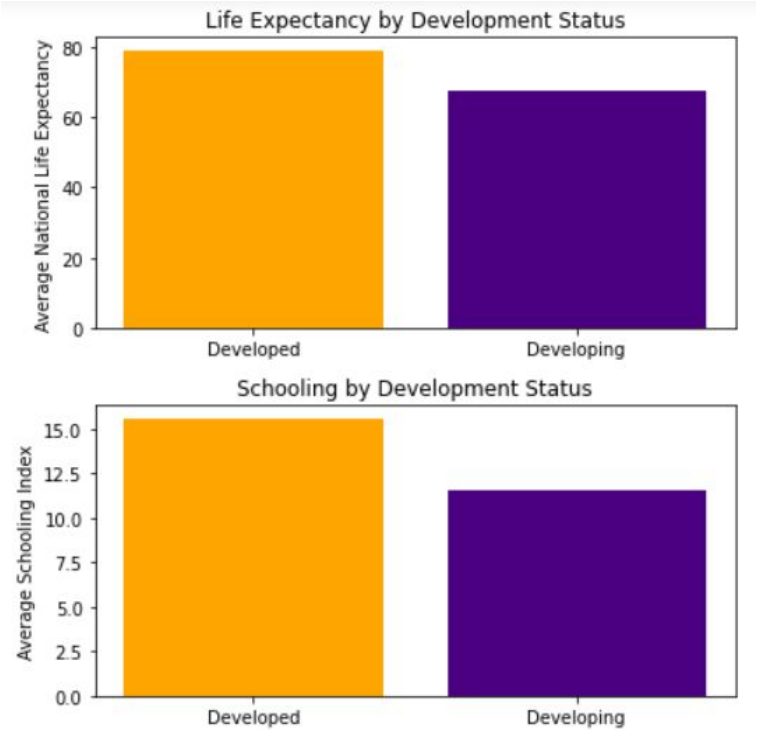
However, we can’t assume that schooling causes higher life expectancy, just because they are correlated.

Schooling and Life Expectancy

Does the development status of a country affect schooling or life expectancy?

Schooling and Life Expectancy

Does the development status of a country affect schooling or life expectancy?



Schooling and Life Expectancy

Does the development status of a country affect schooling or life expectancy?

Interpretation:

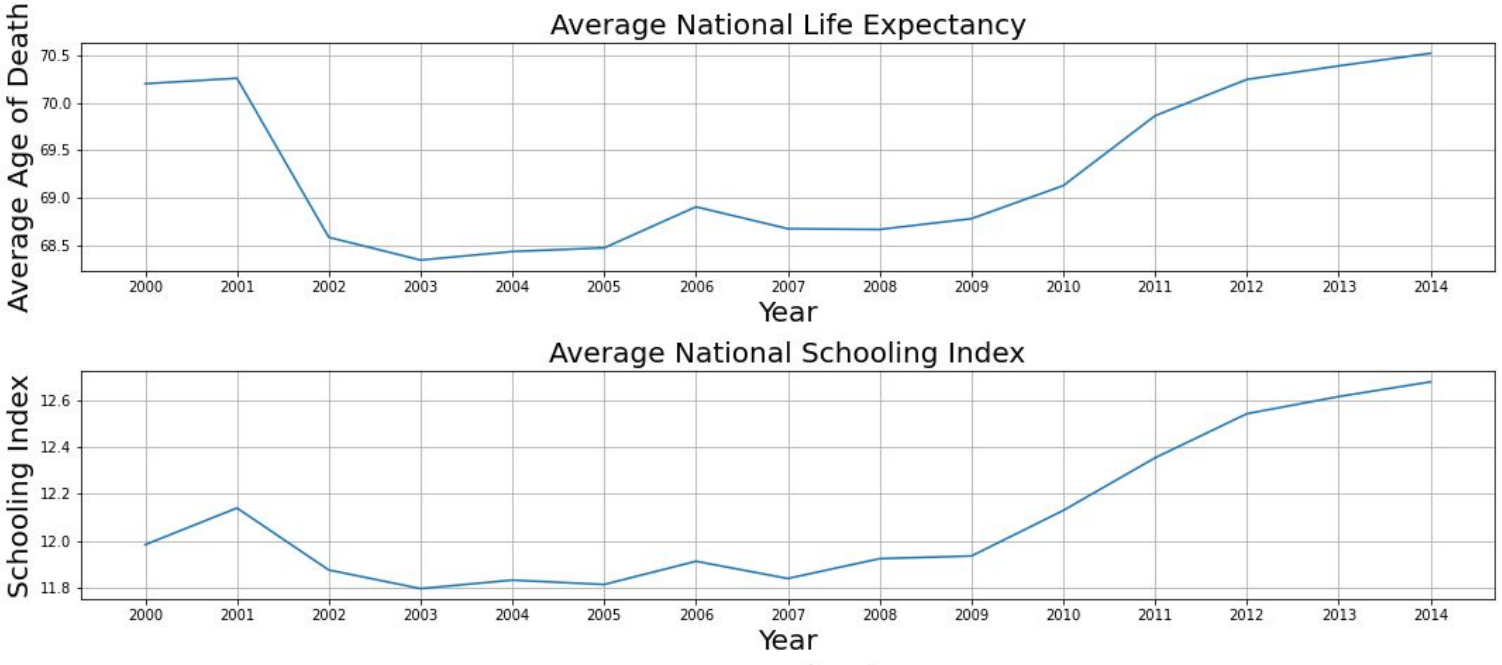
There is a difference in both mean schooling index and life expectancy between developing and developed countries. This may indicate that developed countries either invest more heavily in education, or that education improves the development of countries.

Schooling and Life Expectancy

How have schooling and life expectancy changed over time?

Schooling and Life Expectancy

How have schooling and life expectancy changed over time?



Schooling and Life Expectancy

How have schooling and life expectancy changed over time?

Interpretation:

The life expectancy and schooling both increased between 2000 and 2014. Since schooling index changes seem to follow life expectancy year for year, perhaps schools provide a safe haven for students and reduce deaths.

Recommendation:

Nations should improve access to safe schools in order to raise life expectancy.

Schooling and Life Expectancy

Visualization Source Notebook

[Source Notebook]

Visualization Topics Saved for Next Class

- We are **saving much of the visualization aesthetics concepts** until Thursday.

- <u>Topics saved for next class (if not covered today):</u>
  - Formatting ticks
  - Matplotlib styles/seaborn themes
  - Font customization (titles/axis labels, etc)
  - Figures with multiple subplots
    - Multiple subplots with DIFFERENT figure sizes.
  - Saving/exporting figures (programatically)
  - Interactive visualizations (*if there's time!)*