

CSCE 5320 Scientific Data Visualization

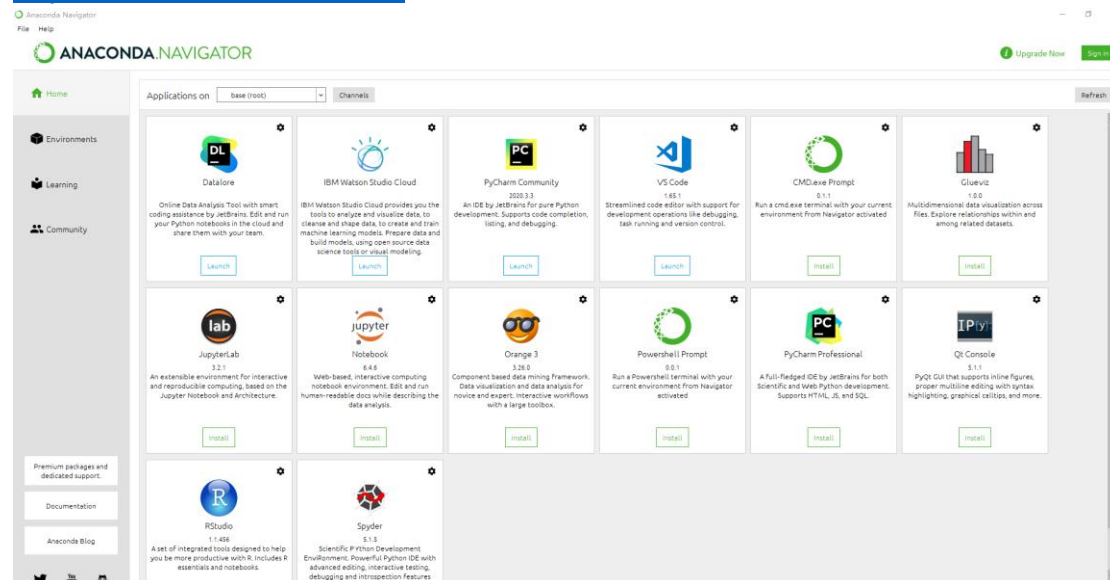
Using Color in Data Visualization

ICE-9: Tutorial

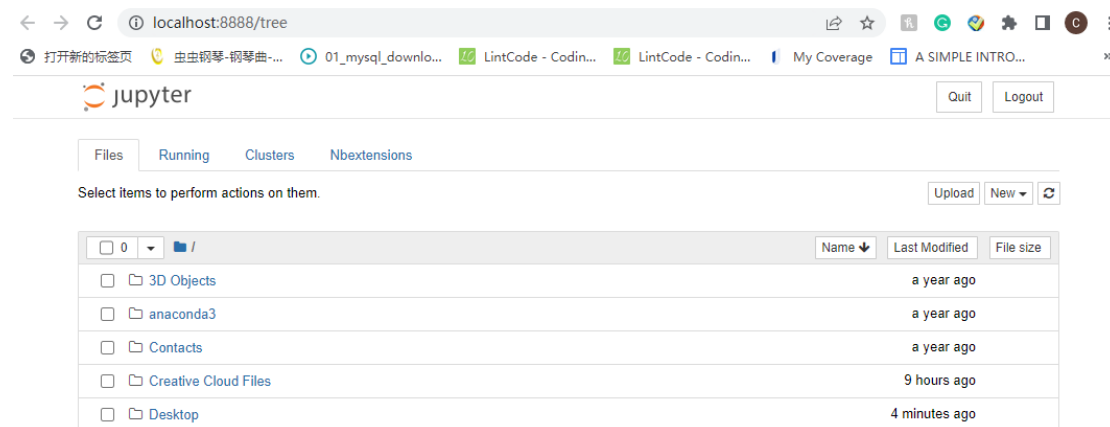
Python Installation:

Anaconda and Jupyter Notebook:

<https://www.anaconda.com/>



<https://jupyter.org/>



You can also use google colab for this lab.

<https://colab.research.google.com/notebooks/charts.ipynb>

● Encoding Data Using Color and Size

Resources for data visualization in Python:

<https://byuidatascience.github.io/python4ds/data-visualisation.html>

Import library in Python for data visualization:

```
import pandas as pd
```

```
import altair as alt
```

Example data: <https://github.com/vega/vega-datasets>

This data shows the global health and population for a number of countries, over the time period of 1955 to 2005.

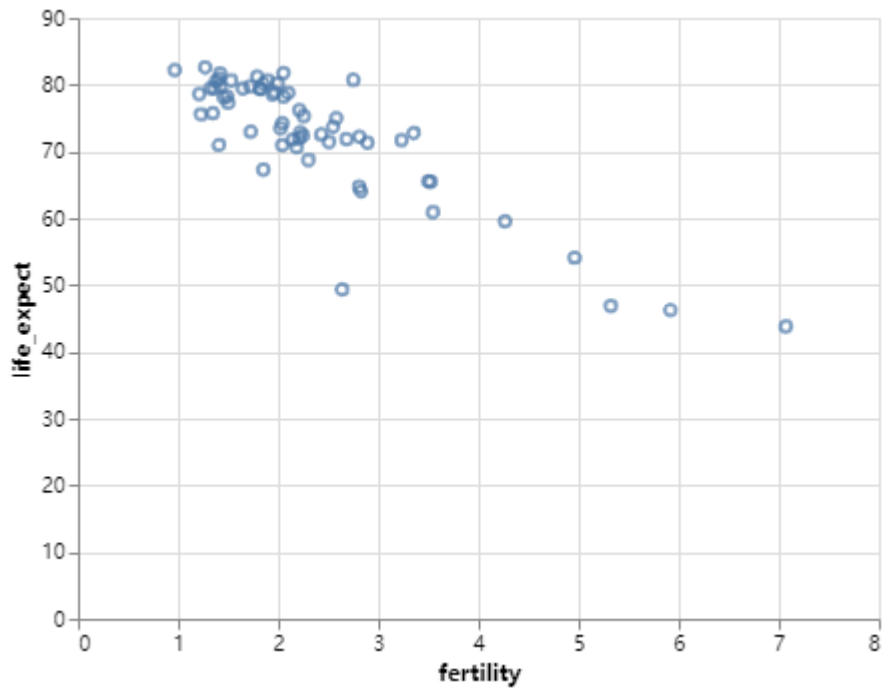
You should use your own data for this lab. The data should contain multiple quantitative values to show different attributes.

Example:

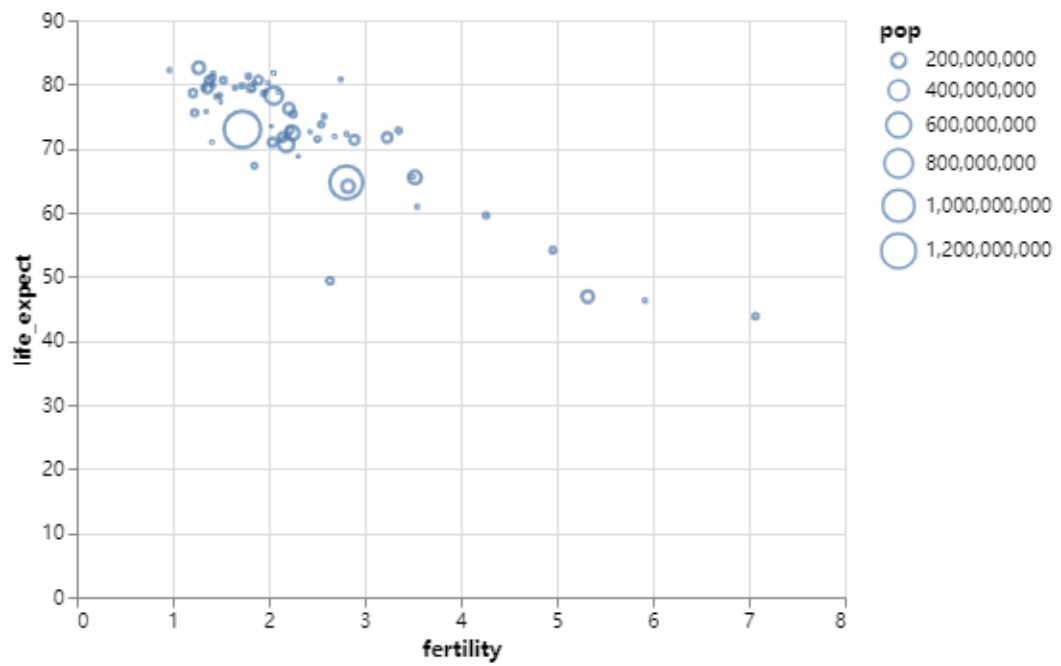
Show part of your data:

	year	country	cluster	pop	life_expect	fertility
10	2005	Afghanistan	0	29928987	43.828	7.0685
21	2005	Argentina	3	39537943	75.320	2.2540
32	2005	Aruba	3	71566	74.239	2.0400
43	2005	Australia	4	20090437	81.235	1.7880
54	2005	Austria	1	8184691	79.829	1.4200
65	2005	Bahamas	3	301790	73.495	2.0221
76	2005	Bangladesh	0	144319628	64.062	2.8260
87	2005	Barbados	3	278870	77.296	1.5000
98	2005	Belgium	1	10364388	79.441	1.6460
109	2005	Bolivia	3	8857870	65.554	3.5000

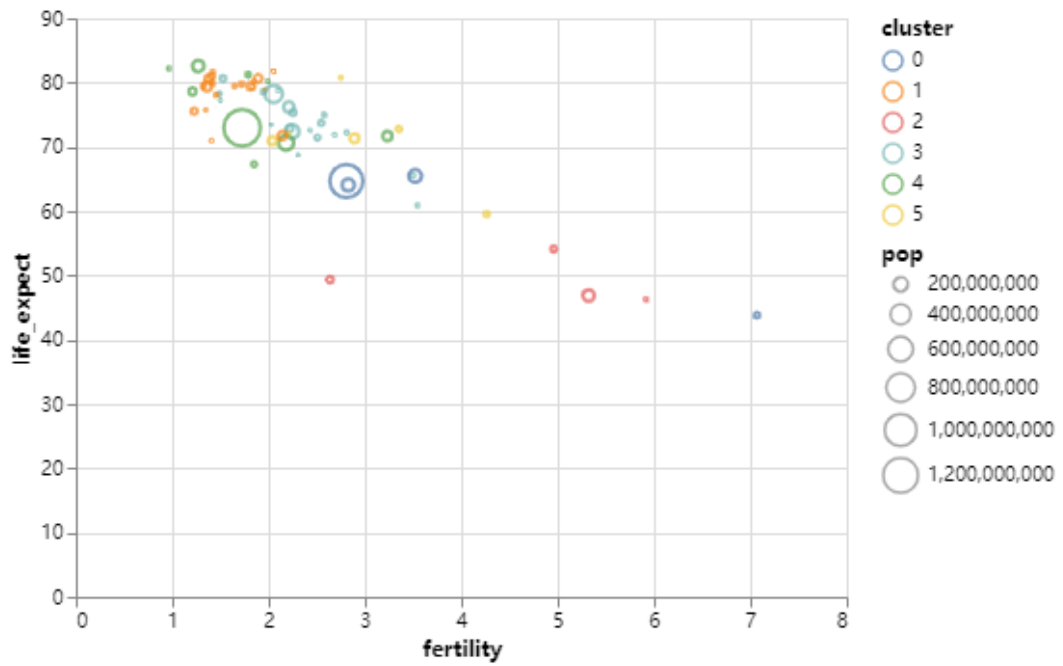
Encoding the data with x-y channels:



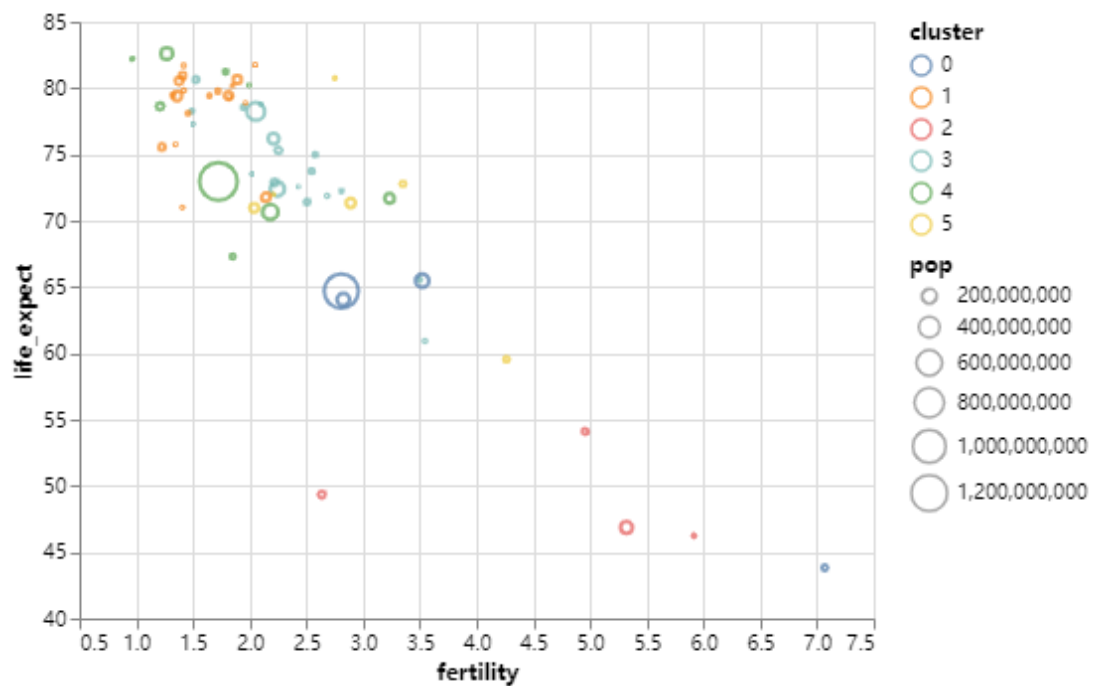
Add Size to your graph:



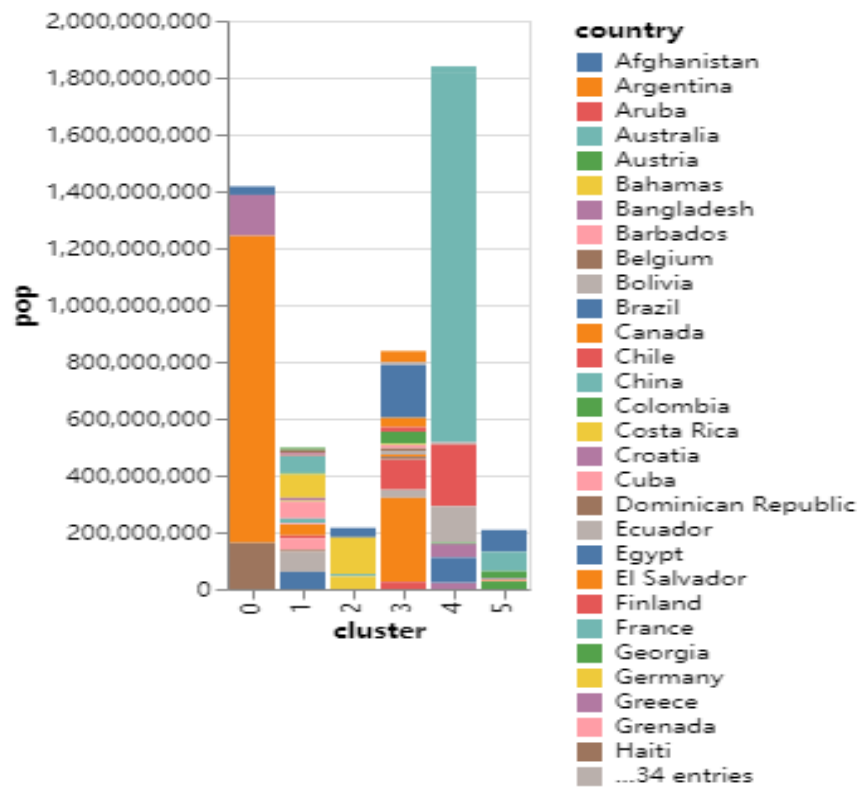
Add Color to your graph:



Optimize your visualization by changing the scale of your graph (remove 0-40 for life-expect in this example):

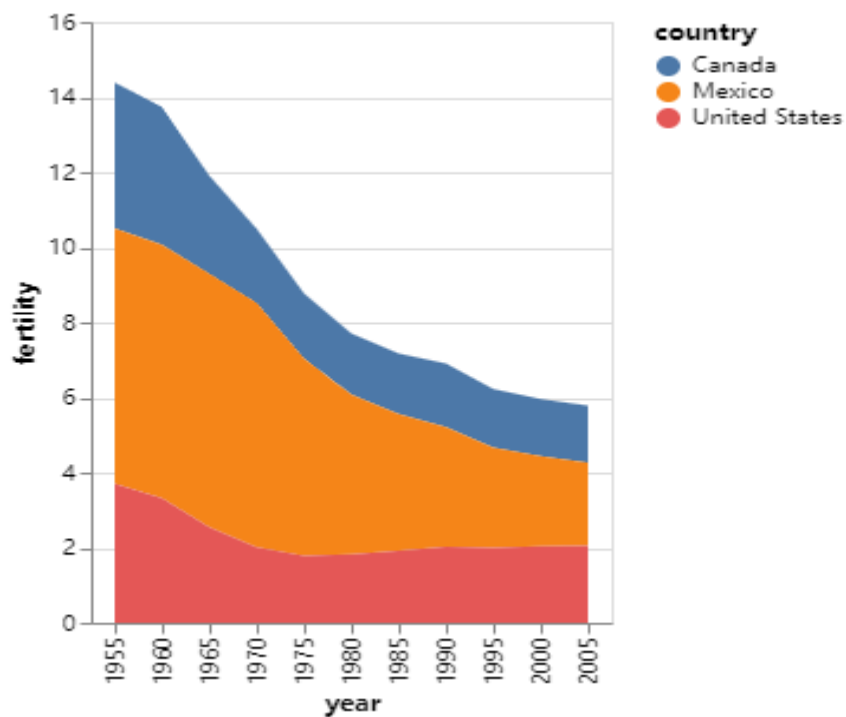


● Stacked & Grouped Bar Chart

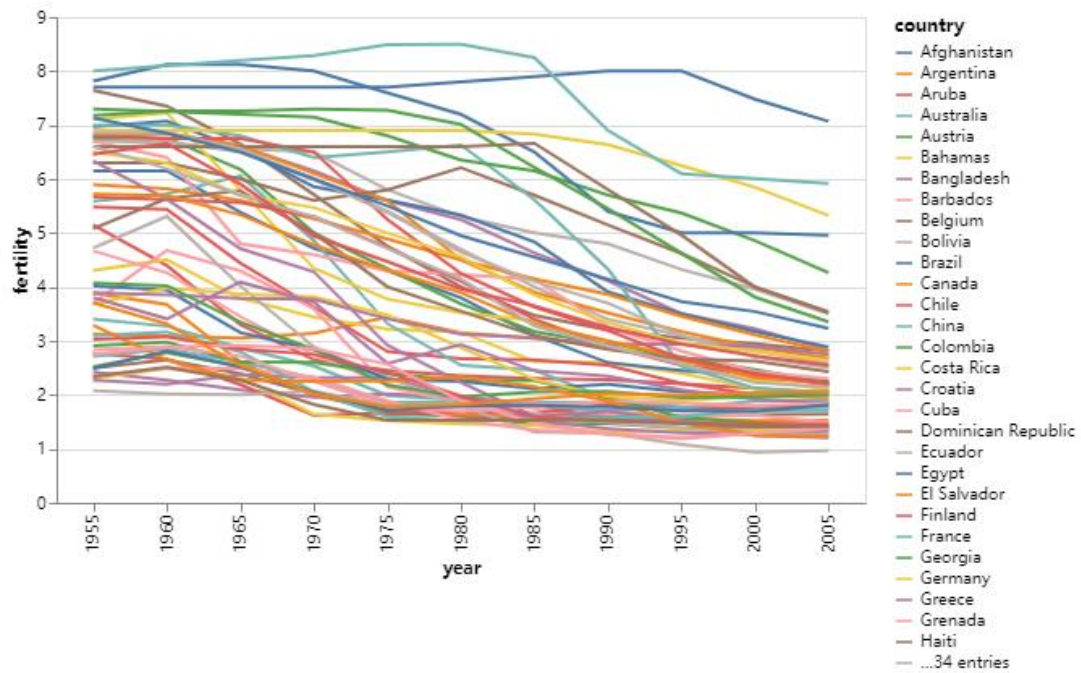


● Stacked Area Chart & Streamgraph

Fertility changes from 1955 to 2005 for the three North American counties.



● Line Chart with Multiple Lines



Fertility Change for countries in East Asia:

