

# Analysis and Visualisation with NetworkX and Altair

## Week 2: Altair

7 December 2020

CDCS Python Libraries Series

with Lucy Havens

# Course Topics

Week 1: NetworkX for network analysis and visualisation

Week 2: Altair for data visualisation

# For Participants

- During class meetings:
  - Take notes
  - Comment and ask questions
  - *Don't worry about writing the code I demo!*
- On your own:
  - Type your own code
  - *Avoid copying and pasting!*

# Getting Set Up

You will need:

- A. Python 3
- B. Pip/pip3 or conda
- C. Altair (along with its complementary packages)
- D. Jupyter Notebooks

# Getting Set Up: Altair

If you're using Anaconda or Miniconda for Jupyter Notebooks, in a Terminal/Shell run:

```
conda install -c conda-forge altair vega_datasets
```

Otherwise, in a Terminal/Shell run one of the following:

```
pip install altair vega_datasets
```

```
pip3 install altair vega_datasets
```

Documentation: <https://altair-viz.github.io/>

# Getting Set Up: Altair for OFFLINE Use

If you're using Jupyter Notebooks, in a Terminal/Shell run:

```
jupyter labextension install @jupyterlab/vega5-extension
```

If you're using Anaconda or Miniconda, in a Terminal/Shell run:

```
conda install vega --channel conda-forge
```

If you're using Jupyter Lab, in a Terminal/Shell run one of the following:

```
pip install vega
```

```
jupyter nbextension install --sys-prefix --py vega
```

Documentation: [https://altair-viz.github.io/user\\_guide/display\\_frontends.html#displaying-charts](https://altair-viz.github.io/user_guide/display_frontends.html#displaying-charts)

# Getting Set Up: Jupyter Notebooks

We'll be using Jupyter Notebooks

3 options:

- A. MyBinder
- B. Using Miniconda or Anaconda
- C. Install to your computer: <https://jupyter.org/install>

# Data Visualization

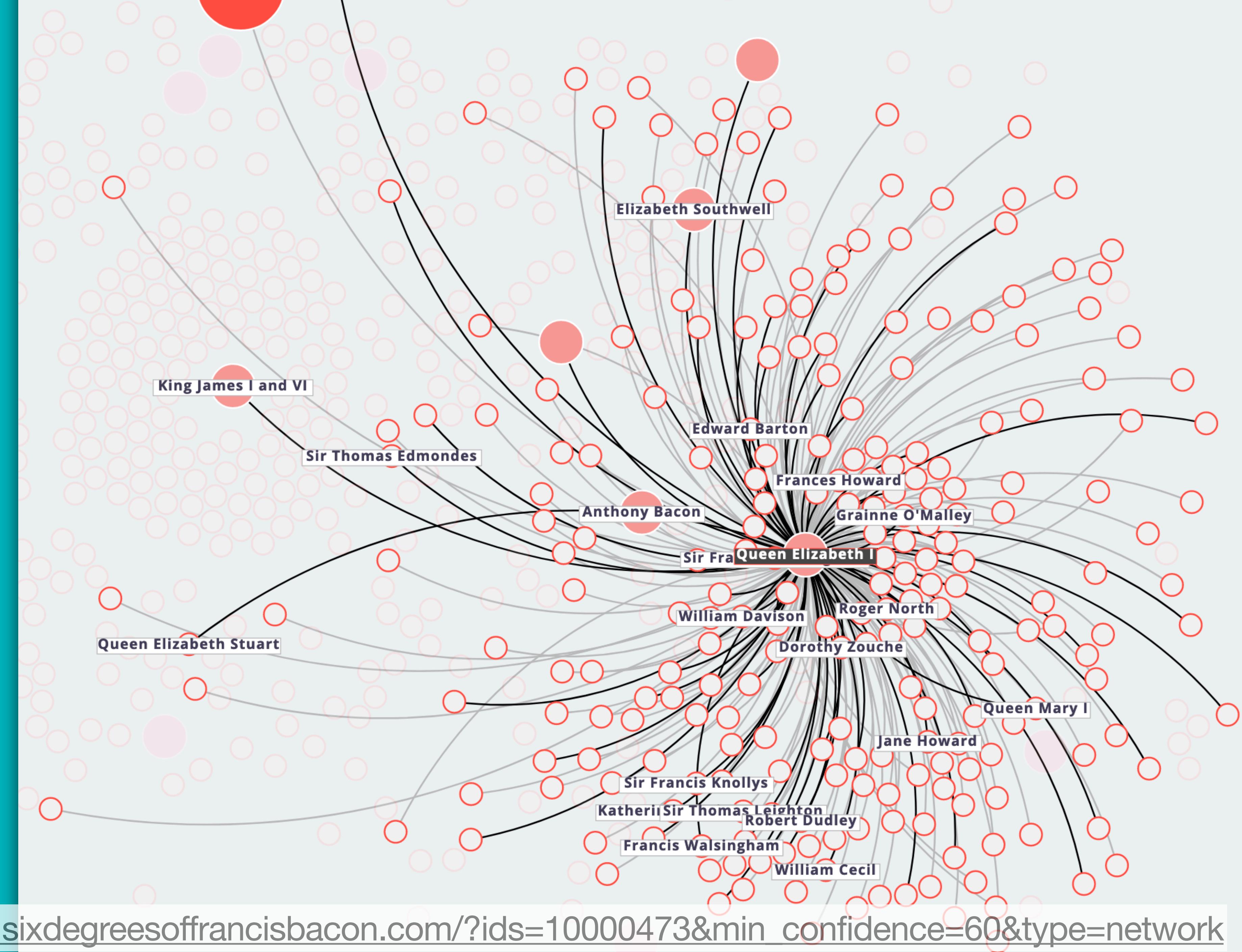
Visually encoding data with:

- Shape
- Color
- Position
- Texture or pattern

Static, dynamic, or interactive

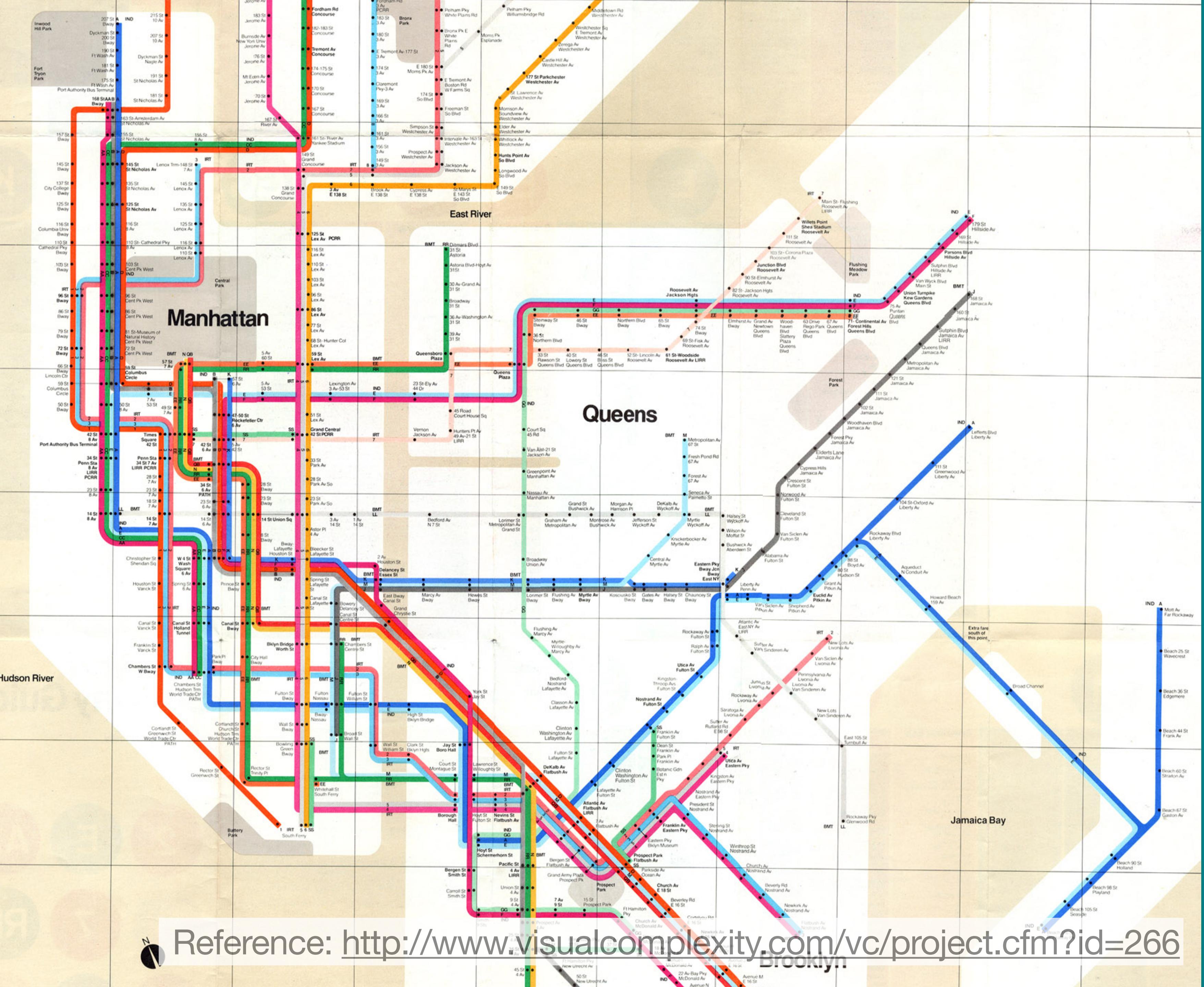
**What types of  
data visualizations  
have you  
seen before?**

# Networks



Reference: [http://www.sixdegreesoffrancisbacon.com/?ids=10000473&min\\_confidence=60&type=network](http://www.sixdegreesoffrancisbacon.com/?ids=10000473&min_confidence=60&type=network)

# Maps



Reference: <http://www.visualcomplexity.com/vc/project.cfm?id=266>

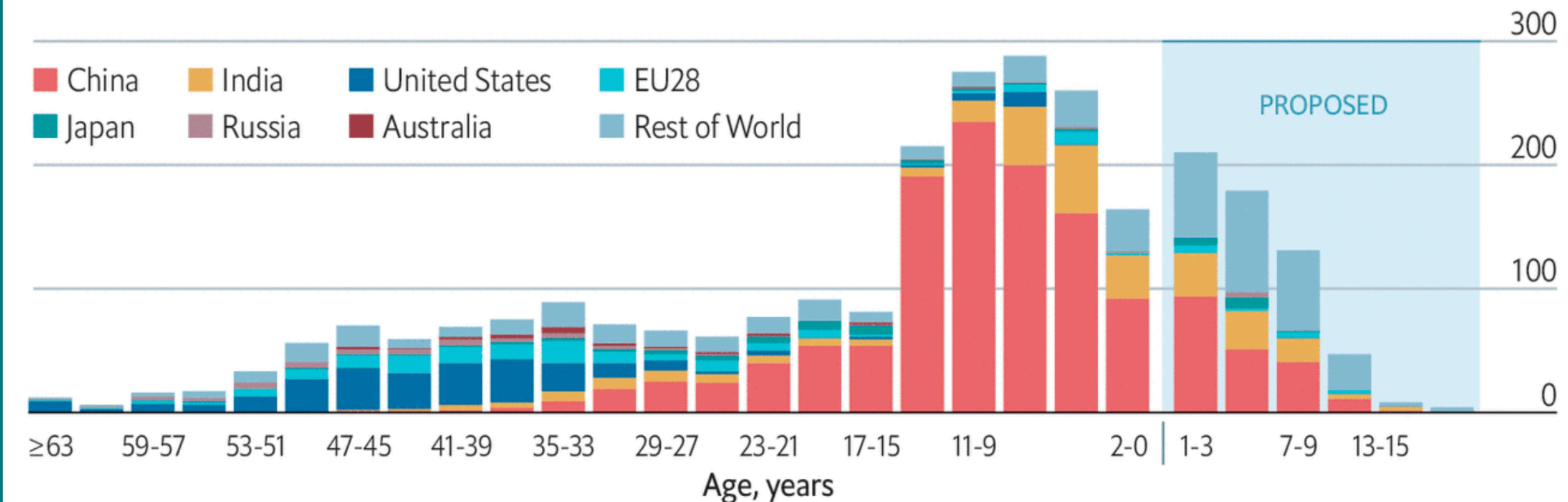
# Bar Charts

## Dethroning King Coal

Digital health care • covid and the economy • China's rise • India's slump

### The young ones

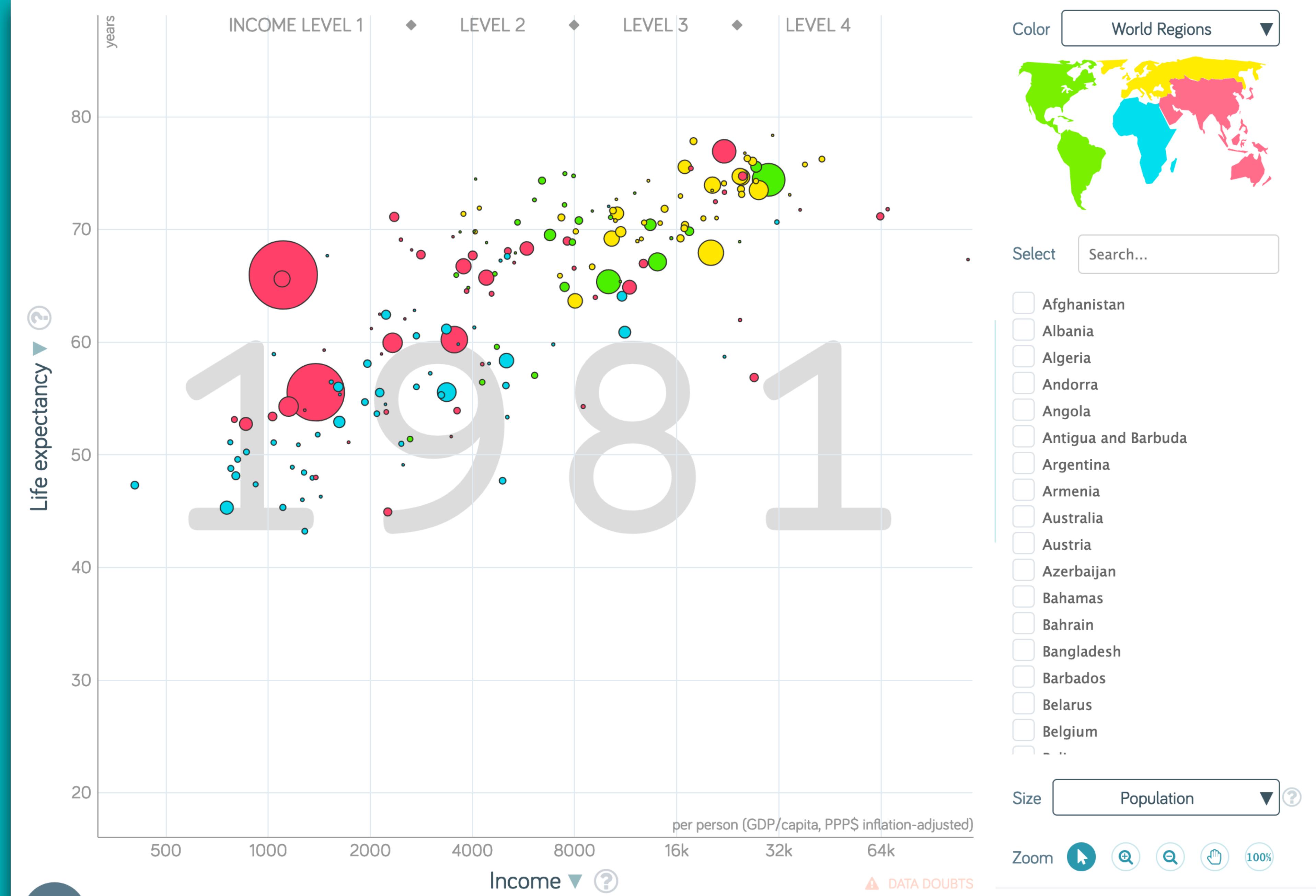
Age structure of electricity-generating coal capacity, 2019, gigawatts



Source: "Committed emissions from existing energy infrastructure jeopardize 1.5°C climate target", by D. Tong et al.

The Economist

# Bubble Charts



Reference: [https://www.gapminder.org/tools/#\\$state\\$time\\$value=1981;&chart-type=bubbles](https://www.gapminder.org/tools/#$state$time$value=1981;&chart-type=bubbles)

**LIVE!**

# More examples

Our World in Data

Articles by topic

Search...

Latest About Donate

All charts Sustainable Development Goals Tracker

Coronavirus pandemic: daily updated research and data. [Read more](#)

Filter interactive charts by title

**Access to Energy**

[Number of people with and without access to clean cooking fuels](#) [Number of people without access to clean fuels for cooking](#)

**Accountability & Transparency**

[Government Transparency Index](#) [Number of countries that adopt policy guarantees for public access to information](#)

**Age Structure**

[Age dependency breakdown by young and old dependents](#) [Age dependency ratio](#)  
[Age dependency ratio projected to 2100](#) [By age group: The growth of the population to 2100](#)  
[Child dependency ratio – the ratio between under-19-year-olds and 20-to-69-year-olds](#) [Median age vs. children per woman](#)  
[Number of children under 5 years old](#) [Old-age dependency ratio](#)  
[Population by age bracket with UN projections](#) [Population by broad age group](#)  
[Population by broad age group projected to 2100](#) [Population growth rate vs. median age](#)  
[Population under five years old](#) [Population younger than 15 years](#)  
[Projected population by age cohort UN \(2015 to 2100\)](#) [Size of young, working age and elderly populations](#)  
[Size of young, working-age and elderly populations](#) [The UN projections of the future population younger than 15 years, by world region](#)  
[The growth of the world population to 2100, by age group](#) [Youth dependency ratio](#)

**Agricultural Employment**

[Agriculture value added per worker](#) [Agriculture value added per worker vs. GDP per capita](#)  
[Average income of small-scale food producers](#) [Employment in agriculture vs GDP per capita](#)  
[GDP per head vs share of agriculture in employment](#) [Number of people employed in agriculture since 1800](#)  
[Share of agricultural landowners who are female](#) [Share of agriculture in GDP vs GDP per capita](#)  
[Share of agriculture in GDP vs GDP per capita](#) [Share of agriculture in total employment](#)  
[Share of employed women working in agriculture vs. GDP per capita](#) [Share of employment in agriculture, industry, and services](#)  
[Share of the labor force employed in agriculture](#)

**Agricultural Production**

[Apple production](#) [Avocado production](#)  
[Banana production](#) [Banana production by region](#)  
[Barley production](#) [Bean production](#)  
[Cashew nut production](#) [Cassava production](#)

[Subscribe](#) [Feedback](#)

Reference: <https://ourworldindata.org/charts>

# Data Visualization: The Data

Data can be:

- **Continuous**
- Discrete
- Nominal
- Ordinal
- Categorical

**Temperature**

-10, -3.8, 0, 0.5, 3, 20

# Data Visualization: The Data

Data can be:

- Continuous
- Discrete
- Nominal
- Ordinal
- Categorical

## Population Size

UK: 66,650,000

US: 328,200,000

World: 7,800,000,000

# Data Visualization: The Data

Data can be:

- Continuous
- Discrete
- Nominal
- Ordinal
- Categorical

## Survey Responses

Strongly Disagree

Disagree

Neither Agree Nor Disagree

Agree

Strongly Agree

# Data Visualization: The Data

Data can be:

- Continuous
- Discrete
- Nominal
- Ordinal
- Categorical

## Race Results

1st place, 2nd place, 3rd  
place, 4th place, ...

# Data Visualization: The Data

Data can be:

- Continuous
- Discrete
- Nominal
- Ordinal
- Categorical

**Favorite Color**

Blue, Orange, Red, Green,  
Yellow, Purple

# Altair

*Altair* is a Python library that works another library, *pandas*

We can make a *Chart* (an Altair object) by specifying:

- **Data source:** table (DataFrame) with the data to visualize
- **Marks:** type of visualization we want to display our data in
- **Encodings:** how we want data to be visually represented

Reference: <https://www.datacamp.com/community/tutorials/altair-in-python>

# **DEMO**

# Altair: Data Types

Data Type	Shorthand Code	Description
quantitative	Q	a continuous real-valued quantity
ordinal	O	a discrete ordered quantity
nominal	N	a discrete unordered category
temporal	T	a time or date value
geojson	G	a geographic shape

Reference: [https://altair-viz.github.io/user\\_guide/encoding.html#encoding-data-types](https://altair-viz.github.io/user_guide/encoding.html#encoding-data-types)

# Altair: Marks

Mark Name	Method	Description	Example
area	<code>mark_area()</code>	A filled area plot.	<a href="#">Simple Stacked Area Chart</a>
bar	<code>mark_bar()</code>	A bar plot.	<a href="#">Simple Bar Chart</a>
circle	<code>mark_circle()</code>	A scatter plot with filled circles.	<a href="#">One Dot Per Zipcode</a>
geoshape	<code>mark_geoshape()</code>	A geographic shape	<a href="#">Choropleth Map</a>
image	<code>mark_image()</code>	A scatter plot with image markers.	<a href="#">Image Mark</a>
line	<code>mark_line()</code>	A line plot.	<a href="#">Simple Line Chart</a>
point	<code>mark_point()</code>	A scatter plot with configurable point shapes.	<a href="#">Multi-panel Scatter Plot with Linked Brushing</a>
rect	<code>mark_rect()</code>	A filled rectangle, used for heatmaps	<a href="#">Simple Heatmap</a>
rule	<code>mark_rule()</code>	A vertical or horizontal line spanning the axis.	<a href="#">Candlestick Chart</a>
square	<code>mark_square()</code>	A scatter plot with filled squares.	N/A
text	<code>mark_text()</code>	A scatter plot with points represented by text.	<a href="#">Bar Chart with Labels</a>
tick	<code>mark_tick()</code>	A vertical or horizontal tick mark.	<a href="#">Simple Strip Plot</a>

Reference: [https://altair-viz.github.io/  
user\\_guide/marks.html](https://altair-viz.github.io/user_guide/marks.html)

# Altair: Encodings

Channel	Altair Class	Description	Example
color	<code>Color</code>	The color of the mark	<a href="#">Simple Heatmap</a>
fill	<code>Fill</code>	The fill for the mark	N/A
fillopacity	<code>FillOpacity</code>	The opacity of the mark's fill	N/A
opacity	<code>Opacity</code>	The opacity of the mark	<a href="#">Horizon Graph</a>
shape	<code>Shape</code>	The shape of the mark	N/A
size	<code>Size</code>	The size of the mark	<a href="#">Table Bubble Plot (Github Punch Card)</a>
stroke	<code>Stroke</code>	The stroke of the mark	N/A
strokeDash	<code>StrokeDash</code>	The stroke dash style	<a href="#">Multi Series Line Chart</a>
strokeOpacity	<code>StrokeOpacity</code>	The opacity of the line	N/A
strokeWidth	<code>StrokeWidth</code>	The width of the line	N/A

Reference: [https://altair-viz.github.io/user\\_guide/encoding.html](https://altair-viz.github.io/user_guide/encoding.html)

# Assignment

Watch a video by Hans Rosling, one of the creators of Gapminder and the author of Factfulness

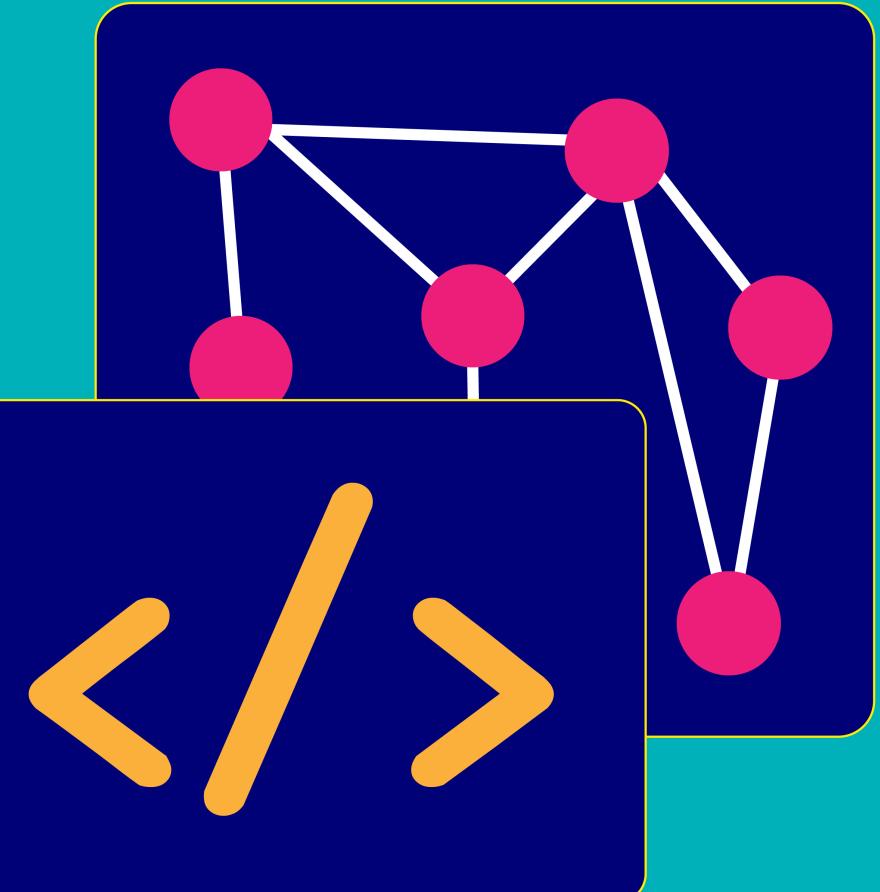
Pick a chart from Our World In Data and answer questions about how it visually encodes information

Create your own charts in Altair

*You can start from the Notebook demoed today!*

# Friday round table on #2!

*Please upload your chosen visualization  
no later than 9 AM GMT on Friday*



# Thanks everyone!

Next course meeting: Friday, 10:00-11:00 AM BST

Office hours available on Wednesday (30 minutes)

*To schedule, please message me on Teams!*