YData: Introduction to Data Science

Class 17: Interactive visualizations and maps

Overview

Interactive visualizations

Mapping

If there is time

- We can work on/discuss projects

Announcement: Homework 7

Homework 7 has been posted!

It is due on Gradescope on Sunday April 2nd at 11pm

- **Be sure to mark each question on Gradescope along with the page that has the answers!**

Reminder: class project

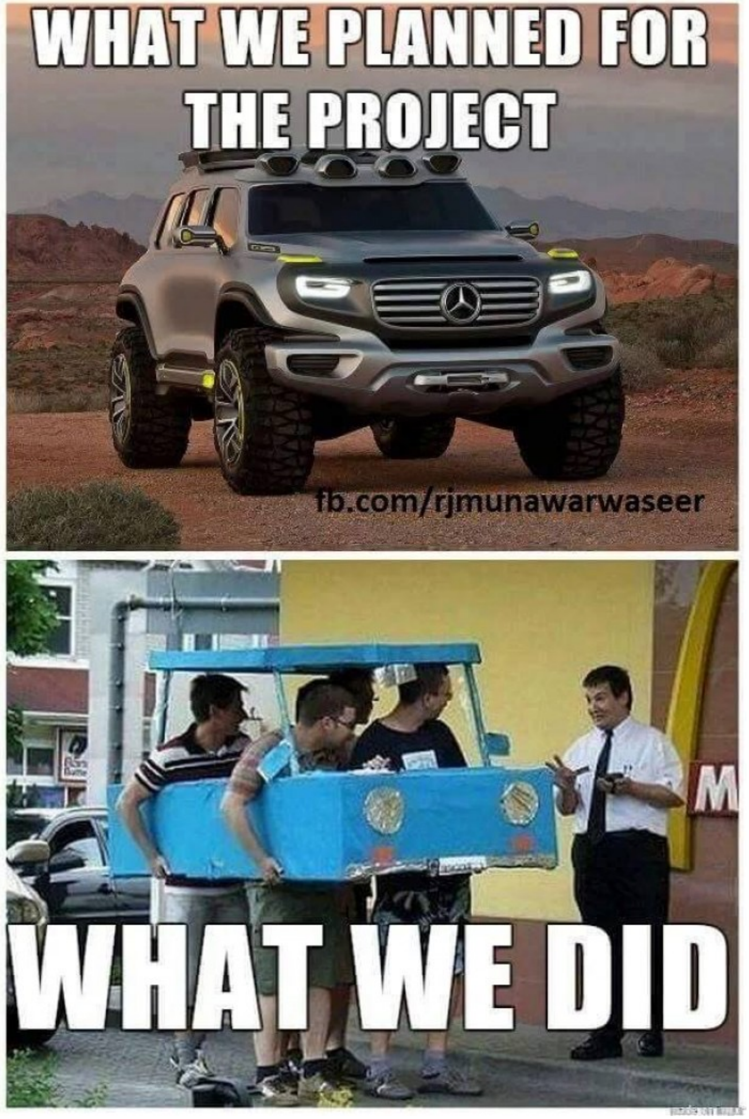
The final project is a 6-10 page Jupyter notebook report where you analyze your own data to address a question that you find interesting

- A project template Jupyter notebooks is on Canvas

A draft of the project is due on April 7th

- ~~So plenty of time, but good to start thinking about it now.~~
- Quickly approaching!

ALL YOU NEED IS MOTIVATION
FALSE: YOU NEED FEAR AND AN APPROACHING DEADLINE
memegenerator.net



Where we are and where we're going…

What we have covered:

- What is Data Science
- Basics of Python (data types, lists, loops)
- Numerical computations (numpy)
- Functions
- Data tables (pandas)
- Data visualization (matplotlib and seaborn)

Today: interactive visualization and mapping

- Connection/reminder of what we have done
- Could be helpful for final project

COMING BACK TO SCHOOL AFTER SPRING BREAK...
I HAVE NO MEMORY OF THIS PLACE...
memegenerator.net

The rest of the semester:

- Statistical analysis
- Machine Learning
- Ethics and conclusions

Interactive visualizations for data exploration

Interactive visualizations are useful for exploring data to find trends

- Visualizations can be shared on the internet
- They can’t be put in static pdfs
  - But can still be useful for your final project to find trends that you can display with static graphics

We will use plotly to create interactive graphics

- import plotly.express as px

Plotly interactive plots

Scatter plots

- px.scatter(data_frame = , x = , y = , size = , color = , hover_name =)

Line plots

- fig = px.line(data_frame =, x =, y = , color = , hover_name = , line_shape =)

Add axis labels

- fig.update_layout(xaxis_title=”X", yaxis_title=”Y")

Let's explore this in Jupyter!

Plotly interactive plots

Sunburst plots

- px.sunburst(data_frame = , path = , values = , color =)

Treemap

- px.treemap(data_frame = , path = , values = , color =)

Let's explore this in Jupyter!

Pivot Tables and heatmaps

Pivot tables aggregate values based on to two grouping variables, and create a table where:

- The columns are the levels of one variable
- The rows are the levels of the other variable

df2 = df.pivot_table(index = "col1", columns = "col2",
values = "col3”, aggfunc = "mean")

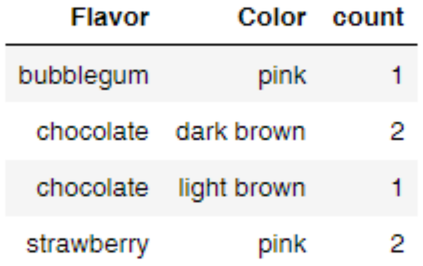
Once we have a 2D table, we can visualize it using:

- px.imshow(df2) # create a heatmap using plotly
- sns.heatmap(df2) # create a heatmap using seaborn

Let's explore this in Jupyter!

Grouping: df.groupby(["col1" col2"]).

col1 col2

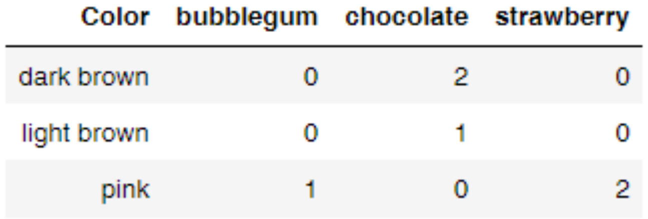


Flavor	Color	count
bubblegum	pink	1
chocolate	dark brown	2
chocolate	light brown	1
strawberry	pink	2

Pivot Table: df.pivot_table()

col1

col2



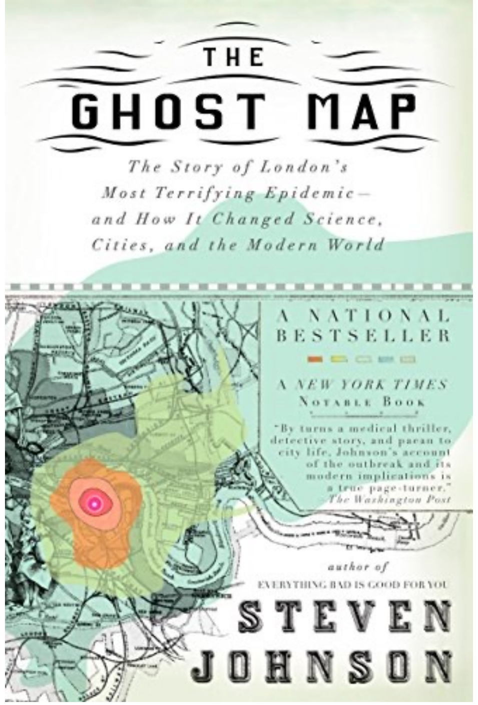
Color	bubblegum	chocolate	strawberry
dark brown	0	2	0
light brown	0	1	0
pink	1	0	2

Maps

Maps to determine the causes of cholera

Visualizing data on a map can be a powerful way to see spatial trends

- One of the first maps used to show spatial trends was created by John Snow to further his case that cholera was a water born illness


THE
GHOST MAP
The Story of London's
Most Terrifying Epidemic—
and How It Changed Science,
Cities, and the Modern World
A NATIONAL BESTSELLER
A NEW YORK TIMES Notable Book
"By turns a medical thriller, detective story, and paean to city life, Johnson's account of the outbreak and its modern implications is a true page-turner."
The Washington Post
author of
EVERYTHING BAD IS GOOD FOR YOU
STEVEN
JOHNSON

THE APPEARANCE AFTER DEATH OF A VICTIM TO THE INDIAN CHOLERA

Cholera in London in the 19th century

Cholera reached London in early 1830s

It was greatly feared as it was often deadly

- An outbreak in 1849 killed over 14,000 people in London

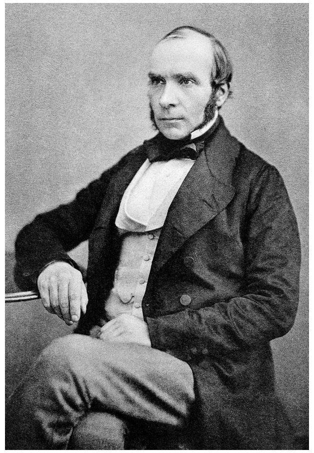
Cause of cholera was unknown. Several theories:

1. Miasmas theory: caused by bad air/smells

- Florence Nightingale, Edwin Chadwick (board of health)

2. Water born disease

- John Snow (anesthesiologist)

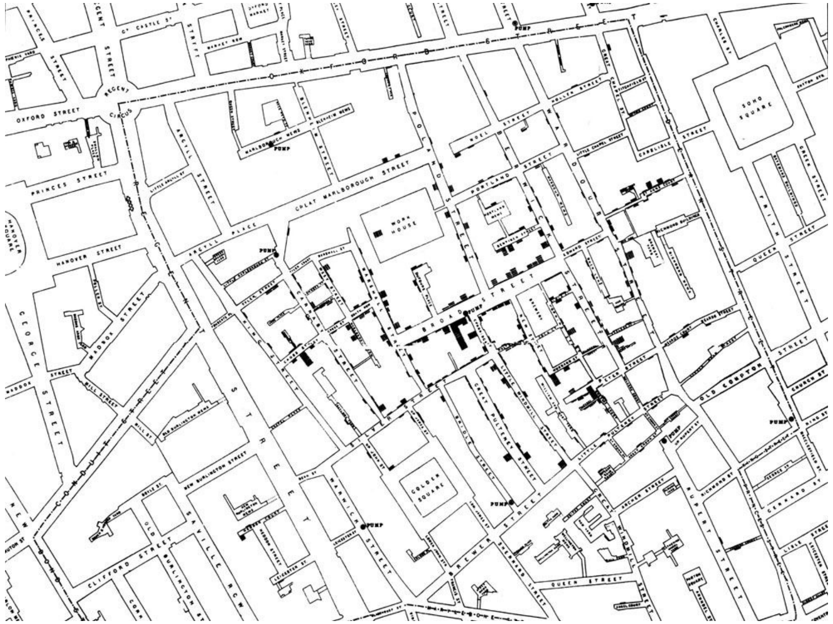


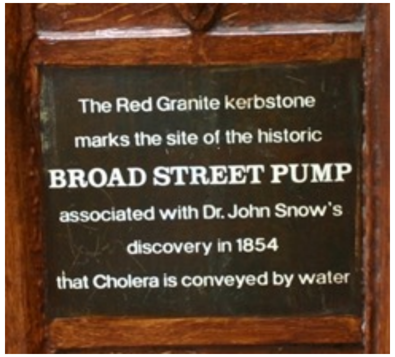
John Snow and spatial mapping

To try to understand the cause of the cholera outbreak of 1854, John Snow plotted a map of cholera deaths

Based on this map and interviews, he concluded that the source of cholera was the Broad Steet well

- He famously removed the handle of the well to prevent the spread of disease
- Now he is considered the founder of epidemiology

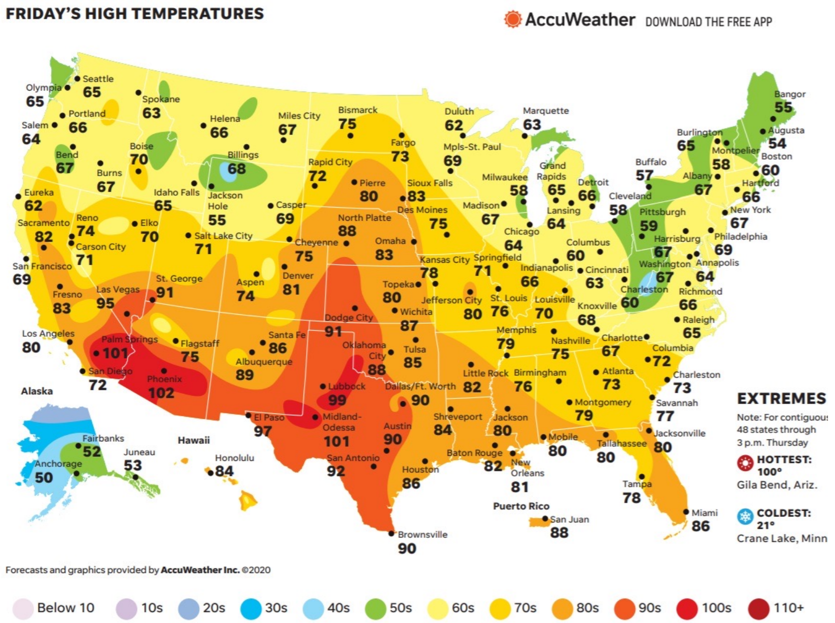




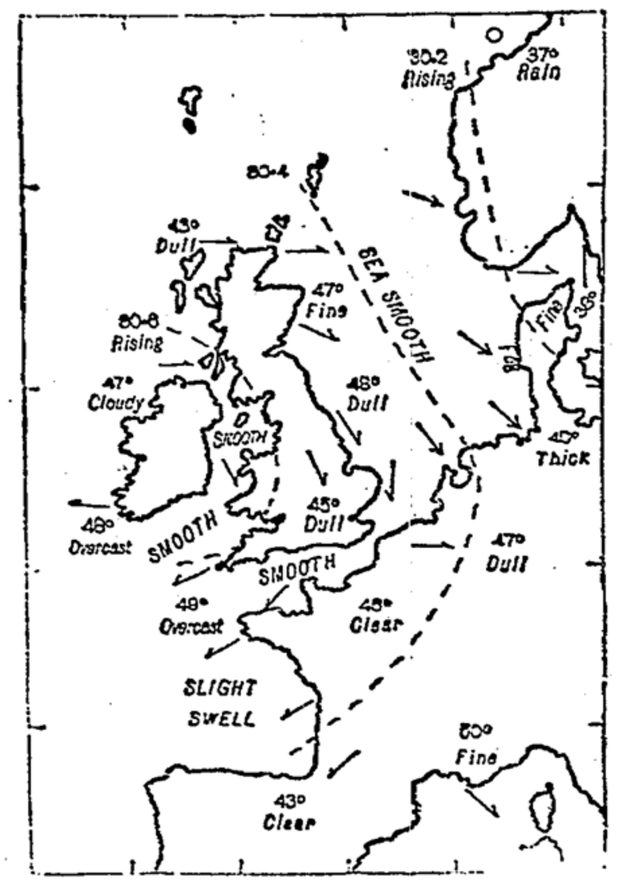
The Red Granite kerbstone
marks the site of the historic
BROAD STREET PUMP
associated with Dr. John Snow's
discovery in 1854
that Cholera is conveyed by water

Maps

Another early use where a map gave insight was the mapping of weather by John Galton in 1875



Weather Chart, March 31, 1875.



The dotted lines indicate the gradations of barometric pressure The variations of the temperature are marked by figures, the state of the sea and sky by descriptive words, and the direction of the wind by arrows—barbed and feathered according to its force. ⊙ denotes calm.

Galton's first weather map (1875)

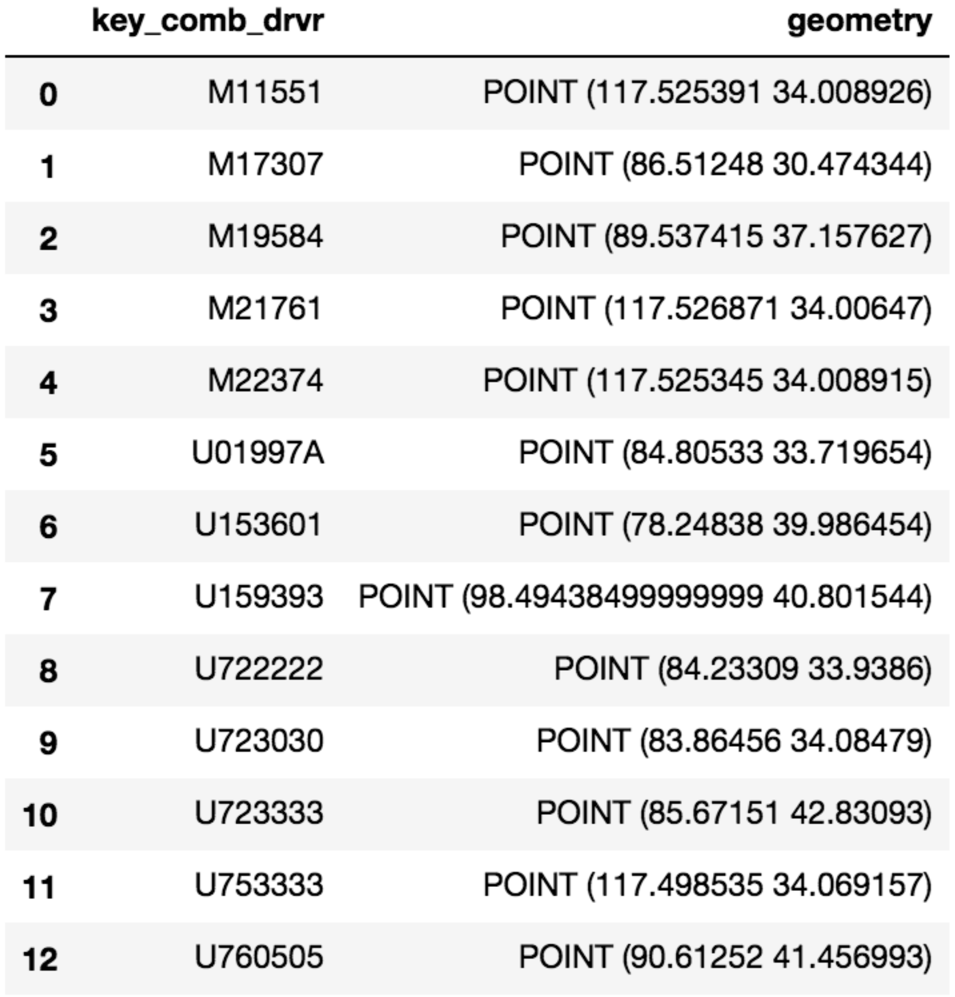
geopandas

To create maps in Python we will use the geopandas package

- import geopandas as gpd

The key object of interest is the geopandas DataFrame

- It is the same as a regular data frame but it has an extra column called ”geometry” that contains geospatial shape features
  - The geometry column as “Shapely” objects used to represent geometric shapes



	key_comb_drvr	geometry
0	M11551	POINT (117.525391 34.008926)
1	M17307	POINT (86.51248 30.474344)
2	M19584	POINT (89.537415 37.157627)
3	M21761	POINT (117.526871 34.00647)
4	M22374	POINT (117.525345 34.008915)
5	U01997A	POINT (84.80533 33.719654)
6	U153601	POINT (78.24838 39.986454)
7	U159393	POINT (98.49438499999999 40.801544)
8	U722222	POINT (84.23309 33.9386)
9	U723030	POINT (83.86456 34.08479)
10	U723333	POINT (85.67151 42.83093)
11	U753333	POINT (117.498535 34.069157)
12	U760505	POINT (90.61252 41.456993)

geopandas

We can read in data as a geopandas DataFrame using

- map = gpd.read_file('my_file.geojson')

We can plot maps using the gpd.plot() function

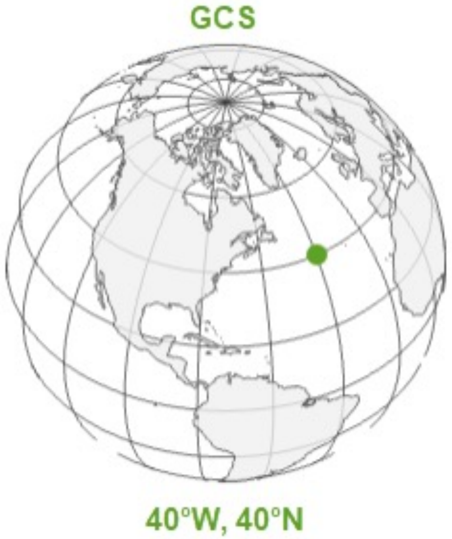
Let's explore this in Jupyter!

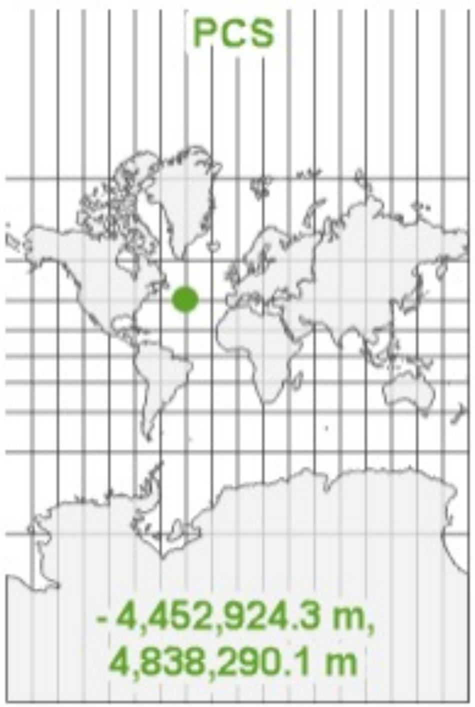
Coordinate reference systems

A coordinate reference system (CRS) is a framework used to precisely measure locations on the surface of the Earth as coordinates

The goal of any coordinate reference system is to create a common reference frame in which locations can be measured precisely and consistently as coordinates, so that any recipient can identify the same location that was originally intended by the originator

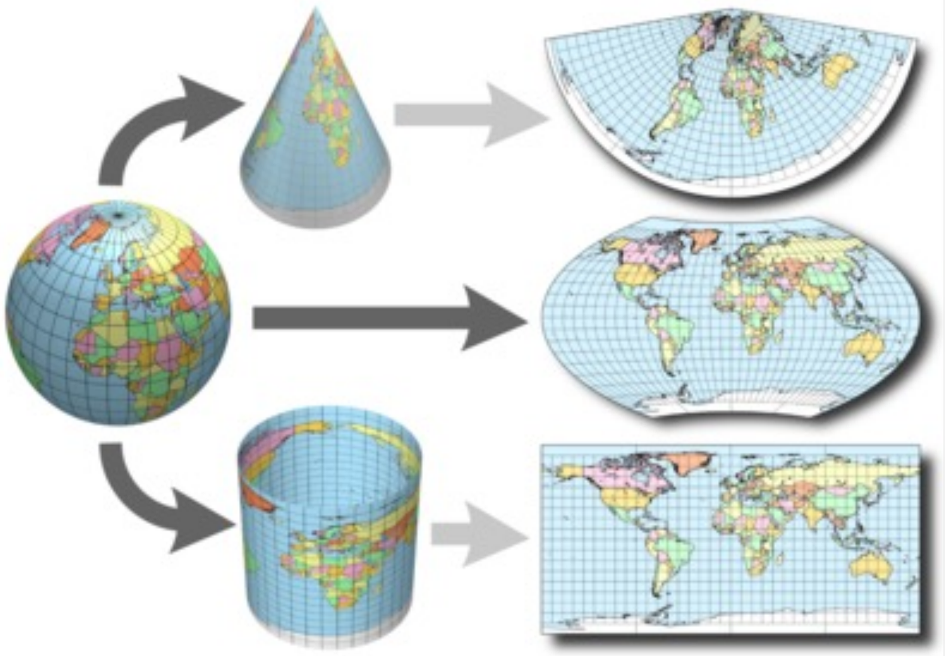
- Needed for aligning different layers on maps





Map projections

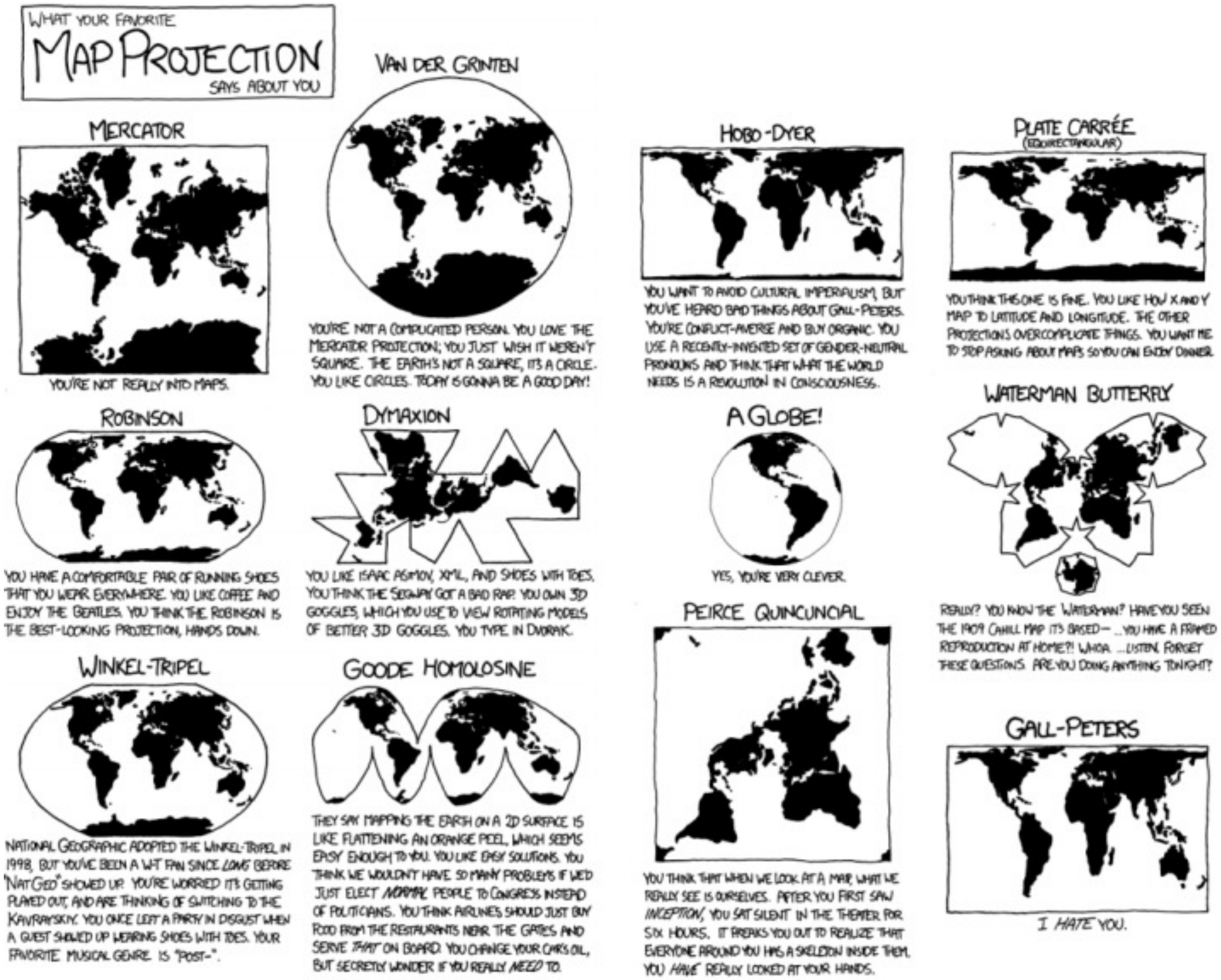
Since the earth is a 3D structure, coordinate systems have to project their data onto a 2D maps



Different projects preserve different properties

- **Mercator projection** keeps angles intact
  - Useful for navigation
- **Eckert IV projection** keeps the size of land areas intact

Let's explore this in Jupyter!

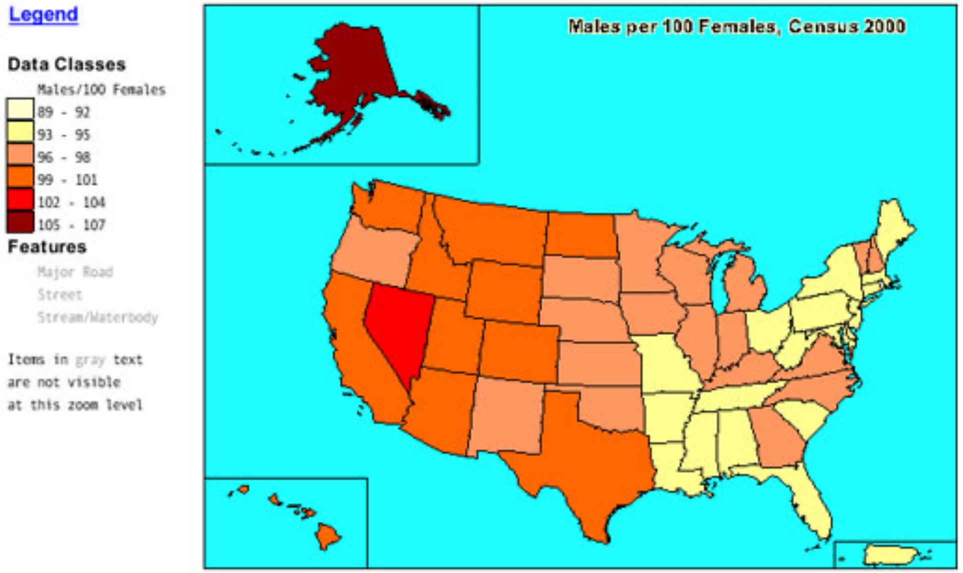

WHAT YOUR FAVORITE MAP PROJECTION SAYS ABOUT YOU
MERCATOR
YOU'RE NOT REALLY INTO MAPS.
VAN DER GRINTEN
YOU'RE NOT A COMPLICATED PERSON. YOU LOVE THE MERCATOR PROJECTION; YOU JUST WISH IT WEREN'T SQUARE. THE EARTH'S NOT A SQUARE, IT'S A CIRCLE. YOU LIKE CIRCLES. TODAY IS GONNA BE A GOOD DAY!
ROBINSON
YOU HAVE A COMFORTABLE PAIR OF RUNNING SHOES THAT YOU WEAR EVERYWHERE. YOU LIKE COFFEE AND ENJOY THE BEATLES. YOU THINK THE ROBINSON IS THE BEST-LOOKING PROJECTION, HANDS DOWN.
DYMAXION
YOU LIKE ISAAC ASIMOV, XML, AND SHOES WITH TOES. YOU THINK THE SEGWAY GOT A BAD RAP. YOU OWN 3D GOGGLES, WHICH YOU USE TO VIEW ROTATING MODELS OF BETTER 3D GOGGLES. YOU TYPE IN DVORAK.
WINKEL-TRIPEL
NATIONAL GEOGRAPHIC ADOPTED THE WINKEL-TRIPEL IN 1998, BUT YOU'VE BEEN A W-T FAN SINCE LONG BEFORE "NAT GEO" SHOWED UP. YOU'RE WORRIED IT'S GETTING PLAYED OUT, AND ARE THINKING OF SWITCHING TO THE KAVRAYSKIY. YOU ONCE LEFT A PARTY IN DISGUST WHEN A GUEST SHOWED UP WEARING SHOES WITH TOES. YOUR FAVORITE MUSICAL GENRE IS "POST-".
GOODE HOMOLOSINE
THEY SAY MAPPING THE EARTH ON A 2D SURFACE IS LIKE FLATTENING AN ORANGE PEEL, WHICH SEEMS EASY ENOUGH TO YOU. YOU LIKE EASY SOLUTIONS. YOU THINK WE WOULDN'T HAVE SO MANY PROBLEMS IF WE'D JUST ELECT NORMAL PEOPLE TO CONGRESS INSTEAD OF POLITICIANS. YOU THINK AIRLINES SHOULD JUST BUY FOOD FROM THE RESTAURANTS NEAR THE GATES AND SERVE THAT ON BOARD. YOU CHANGE YOUR CAR'S OIL, BUT SECRETLY WONDER IF YOU REALLY NEED TO.
HOBO-DYER
YOU WANT TO AVOID CULTURAL IMPERIALISM, BUT YOU'VE HEARD BAD THINGS ABOUT GALL-PETERS. YOU'RE CONFLICT-AVERSE AND BUY ORGANIC. YOU USE A RECENTLY-INVENTED SET OF GENDER-NEUTRAL PRONOUNS AND THINK THAT WHAT THE WORLD NEEDS IS A REVOLUTION IN CONSCIOUSNESS.
A GLOBE!
YES, YOU'RE VERY CLEVER.
PEIRCE QUINCUNCIAL
YOU THINK THAT WHEN WE LOOK AT A MAP, WHAT WE REALLY SEE IS OURSELVES. AFTER YOU FIRST SAW INCEPTION, YOU SAT SILENT IN THE THEATER FOR SIX HOURS. IT FREAKS YOU OUT TO REALIZE THAT EVERYONE AROUND YOU HAS A SKELETON INSIDE THEM. YOU HAVE REALLY LOOKED AT YOUR HANDS.
PLATE CARRÉE
(EQUIRECTANGULAR)
YOU THINK THIS ONE IS FINE. YOU LIKE HOW X AND Y MAP TO LATITUDE AND LONGITUDE. THE OTHER PROJECTIONS OVERCOMPLICATE THINGS. YOU WANT ME TO STOP ASKING ABOUT MAPS SO YOU CAN ENJOY DINNER.
WATERMAN BUTTERFLY
REALLY? YOU KNOW THE WATERMAN? HAVE YOU SEEN THE 1909 CAHILL MAP IT'S BASED— ...YOU HAVE A FRAMED REPRODUCTION AT HOME?! WHOA. ...LISTEN, FORGET THESE QUESTIONS. ARE YOU DOING ANYTHING TONIGHT?
GALL-PETERS
I HATE YOU.

Maps

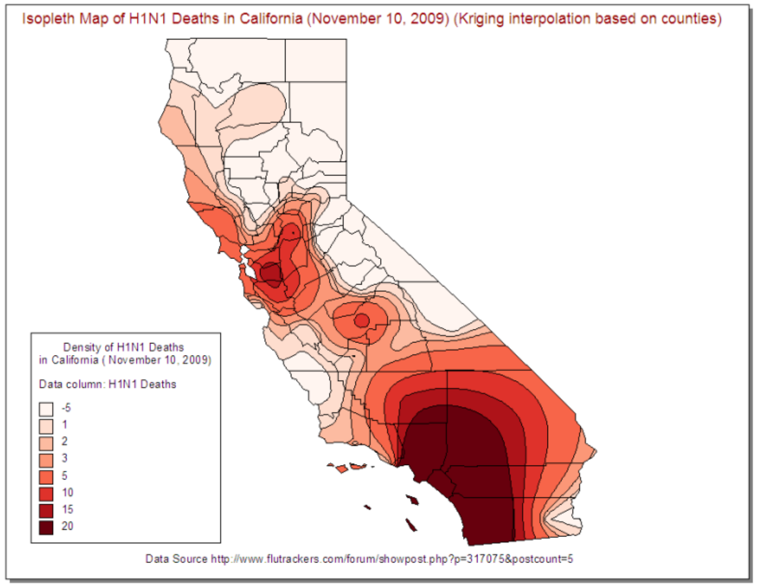
Choropleth maps: shades/colors in predefined areas based on properties of a variable

Isopleth maps: creates regions based on constant values

Choropleth map



Isopleth map


Isopleth Map of H1N1 Deaths in California (November 10, 2009) (Kriging interpolation based on counties)
Density of H1N1 Deaths
in California (November 10, 2009)
Data column: H1N1 Deaths
-5
1
2
3
5
10
15
20
Data Source http://www.flutrackers.com/forum/showpost.php?p=317075&postcount=5

Choropleth maps

We can create choropleth maps using geopandas by joining region information on to a geopandas DataFrame that has a map

We can then use the gpd.plot() method to visualize the map

Let's explore this in Jupyter!