

JSC370
A4 - SWL/NUMINA PROJECT
Prof. Nathan Taback

A4 QUESTIONS

- How can we use the data to understand dwell time (i.e., where people spend a lot of time) at 307? Which areas of 307 do people pass through? Where do people tend to linger? How does dwell time change over time?
- During events at 307 do people tend to dwell in certain spots or choose certain desire lines(i.e., where people come from and where they go)? Are dwell times or desire lines different compared to times when there are no events?
- 307 requires maintenance after 500 visitors or 500 hours of use, when should 307 plan on scheduling maintenance operations (e.g., what days and times)?

DATA AND PRIVACY

- Are you selling your data to Numina when it records you?
- More generally, do you own your own data?
- If you do own your own data then do you have the right to keep it private even if it can't be used to reidentify you?

The screenshot shows the Numina website's homepage. At the top, there is a navigation bar with links for Why, Solution, Blog, Team, and a green link for "We're hiring". To the right of the navigation bar are "Login" and "Request demo" buttons. Below the navigation, there is a section with text about goals in urban planning, traffic safety, and other important fields. It mentions a video that created an opportunity for further, unintended data extraction later. A list of theses and tenets follows, emphasizing privacy and security by design. The Privacy Policy is embedded below, stating that true privacy protection involves not just *not* identifying people; it must also prevent re-identifiability. This statement is highlighted with a red border. The text also notes that they can do this by avoiding excessive data collection and aggregating data so that individual behaviors are not visible.

On Personal Data,
Forgiveness, and
The "Right to Be Forgotten"

Luciano Floridi

INTERACTION IN DATA VISUALIZATION

Interaction within the data and information visualization context is a mechanism for modifying what the users see and how they see it.
(Ward, Interactive Data Visualization)

INTERACTION TECHNIQUES

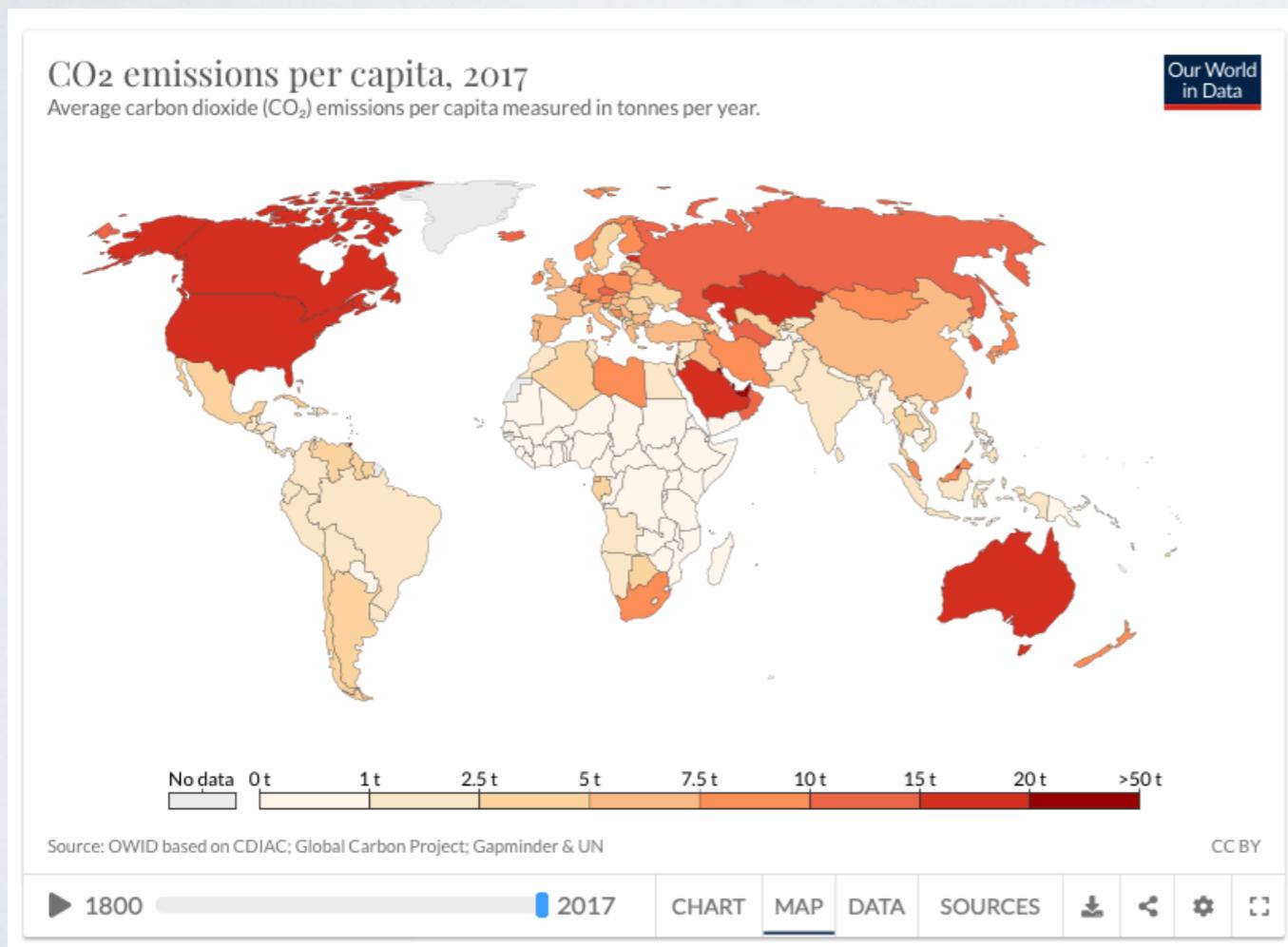
- **Navigation:** user controls for altering the position of the camera and for scaling the view (what gets mapped to the screen) such as panning, rotating, and zooming.
- **Selection:** user controls for identifying an object, a collection of objects, or regions of interest to be the subject of some operation, such as highlighting, deleting, and modifying.
- **Filtering:** user controls for reducing the size of the data being mapped to the screen, either by eliminating records, dimensions, or both.
- **Sort:** user controls for changing the way data is mapped to graphical entities or attributes, such as reordering the data or layouts, thereby providing a different way of viewing a data subset.
- **Derive:** transform variables.
- **Hybrid:** user controls combining several of the above in one technique, for example, increasing the screen space assigned to one or more focus areas to enable users to see details, while showing the other areas of data in a smaller space, in a way that preserves context

Which interaction techniques are used?

What is the message?

Who is the audience?

What does the audience need to know?



<https://ourworldindata.org/per-capita-co2>

Which interaction techniques are used?

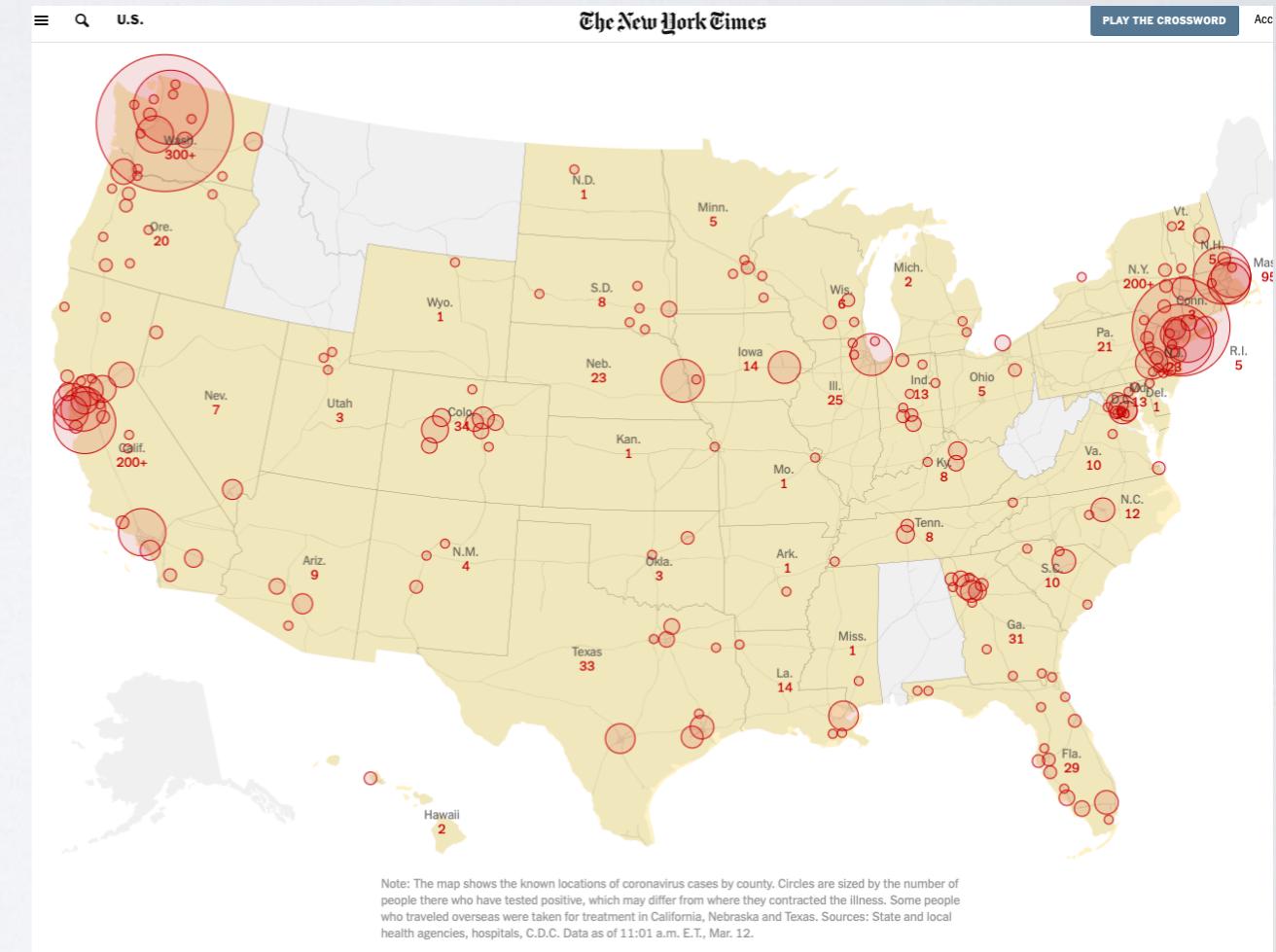
What is the message?

Who is the audience?

What does the audience need to know?



<https://coronavirus.jhu.edu/map.html>

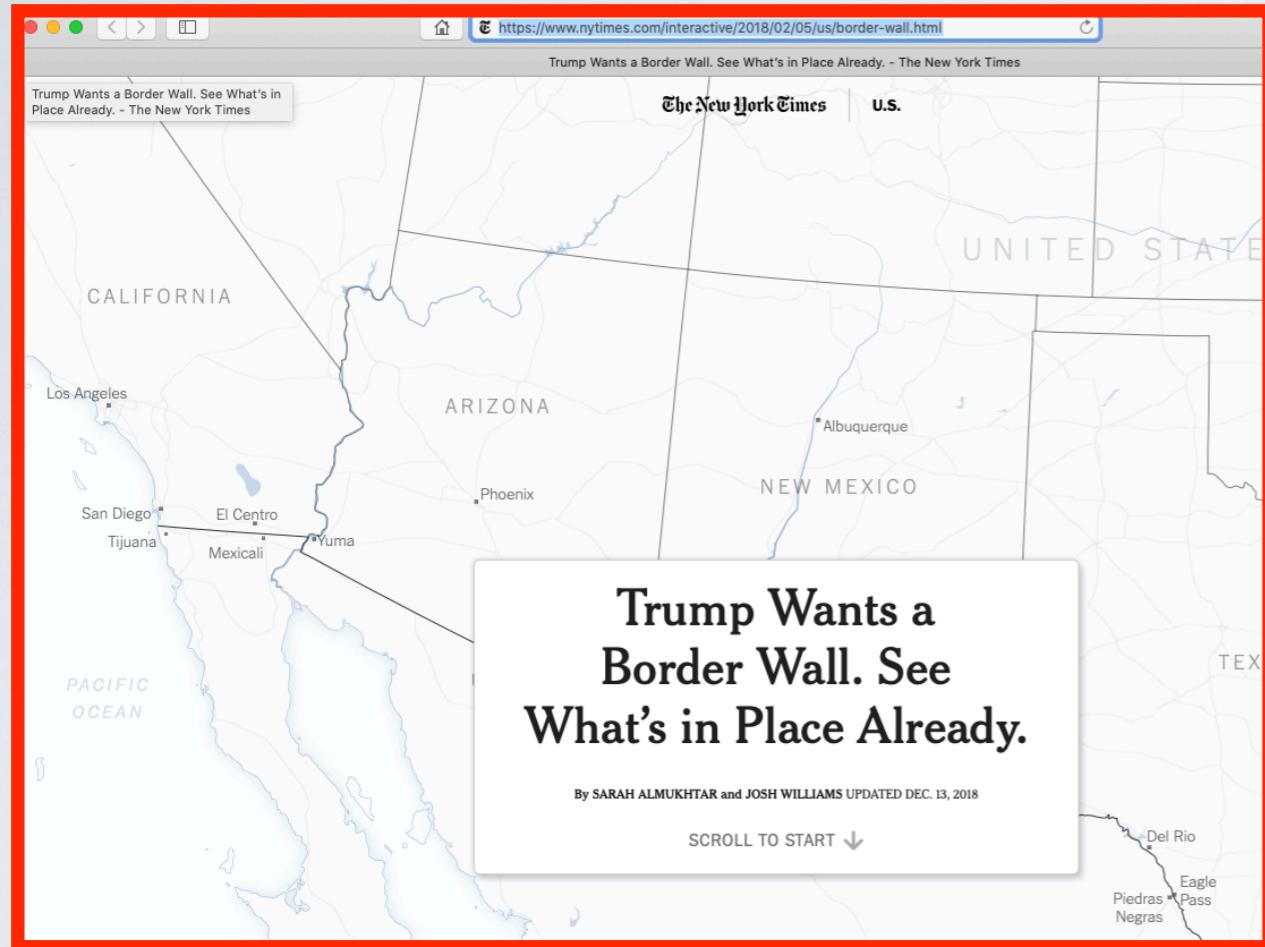


<https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html>

TELLING STORIES WITH DATA

- Data are represented by numbers and words.
- Data is a representation of something in real life.
- Statistics and visualization can help tell a story.
- Data scientists can play a large role in deciding how to tell the story, and what story to tell.

TELLING STORIES WITH DATA



Parallax scrolling is a web site trend where the background content (i.e. an image) is moved at a different speed than the foreground content while scrolling. ([Ref: w3schools.com](https://www.w3schools.com))

<https://www.nytimes.com/interactive/2018/02/05/us/border-wall.html>

Which interaction techniques are used?
What is the story?
Is parallax scrolling effective in telling the story?

- This graphic presents data in a clear and concise manner.
- Important points, areas are annotated, symbols and colours explained, and it's easy to see the story in the data.
- This is a simple line chart, but design elements help tell a better story.
- Line width and colour direct your eyes to what's important.

graph¹ | gra:f, graf |

noun

a diagram showing the relation between variable quantities, typically of two variables, each measured along one of a pair of axes at right angles.

- *Mathematics* a collection of points whose coordinates satisfy a given relation.

verb [with object]

plot or trace on a graph.

graphic | 'grafɪk |

adjective

- 1 relating to visual art, especially involving drawing, engraving, or lettering: *his mature graphic work*.
• *Computing* relating to or denoting a visual image: *graphic information such as charts and diagrams*.

- 2 giving clear and vividly explicit details: *a graphic account of the riots*.

- 3 of or in the form of a graph.

- 4 [attributive] *Geology* of or denoting rocks having a surface texture resembling cuneiform writing.

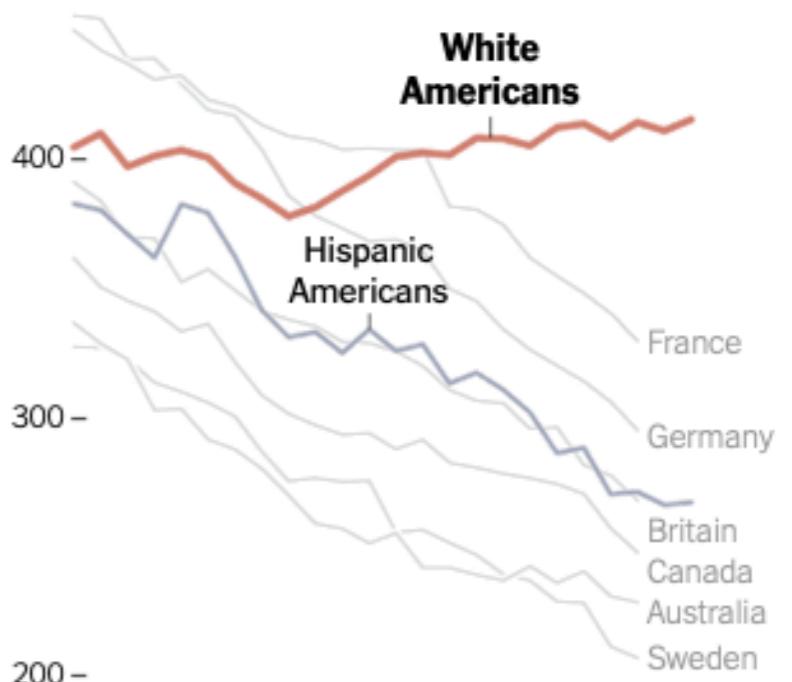
noun *Computing*

a graphical item displayed on a screen or stored as data.

Dying in Middle Age

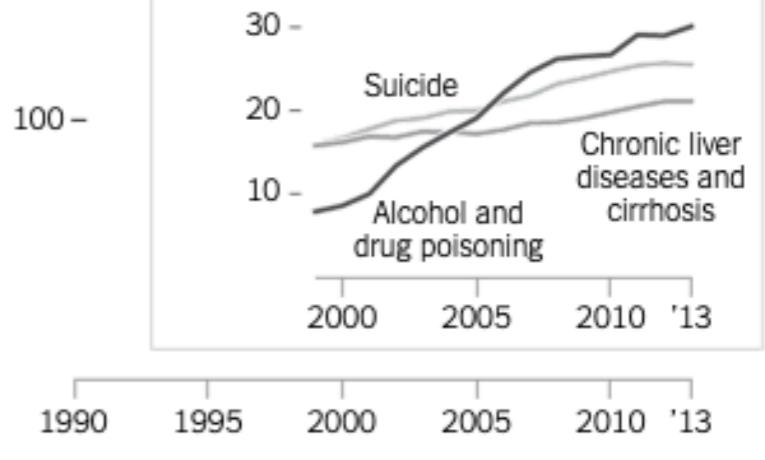
Death rates are rising for middle-aged white Americans, while declining in other wealthy countries and among other races and ethnicities. The rise appears to be driven by suicide, drugs and alcohol abuse.

DEATHS per 100,000 people aged 45–54



INCREASING CAUSES OF DEATHS

Per 100,000 white Americans, 45–54



Sources: Anne Case and Angus Deaton; PNAS

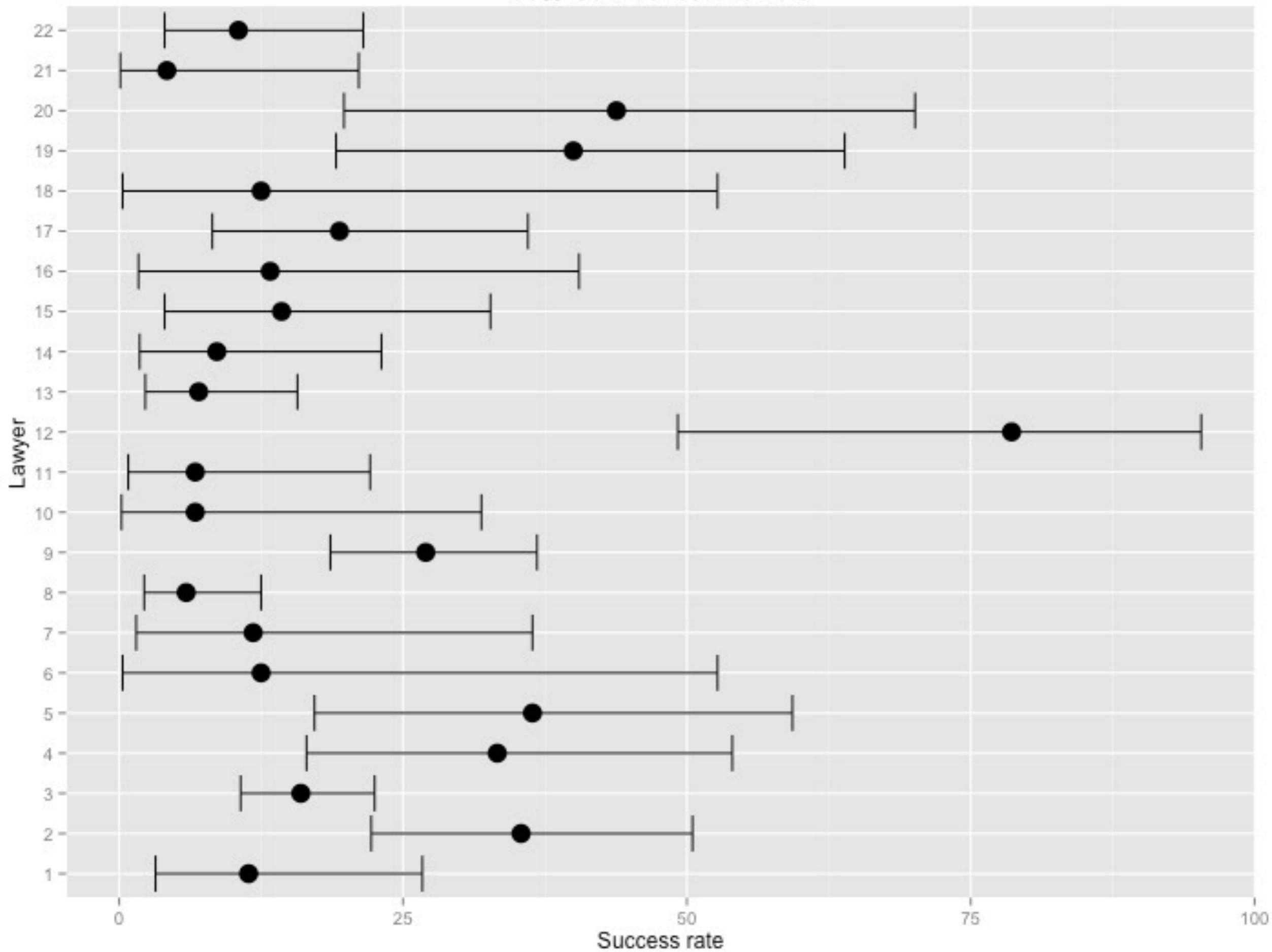
By The New York Times

Counsel	Abandoned/ Withdrawn	Negative	Positive	Total	Recognition Rate	95% Lower Limit	95% Upper Limit
MANZARARU, LEONARD	22	3	11	36	78.6	49.2	95.3
VALOIS, STEPHANIE	24	9	7	40	43.8	19.8	70.1
Vallieres, Alain	20	12	8	40	40	19.1	63.9
GOLDMAN, JEFFREY	9	14	8	31	36.4	17.2	59.3
BHATTI, ROGER	66	31	17	114	35.4	22.2	50.5
FINE, DANIEL	70	18	9	97	33.3	16.5	54
Ivanyi, Peter	149	73	27	249	27	18.6	36.8
SILCOFF, MAUREEN	16	29	7	52	19.4	8.2	36
FARKAS, JOSEPH	223	137	26	386	16	10.7	22.5
Rodrigues, Roger	24	24	4	52	14.3	4	32.7
SARKOZI, JOZEF	10	13	2	25	13.3	1.7	40.5
GRICE, JOHN	26	7	1	34	12.5	0.3	52.7
TAHERI, DJAWID	65	7	1	73	12.5	0.3	52.7
HEGYI, ILDIKO	14	15	2	31	11.8	1.5	36.4
	297	31	4	332	11.4	3.2	26.7
YOUNES, DIANA	18	51	6	75	10.5	4	21.5
NO COUNSEL, IDENTIFIED	67	32	3	102	8.6	1.8	23.1
NO COUNSEL,	66	66	5	137	7	2.3	15.7
JASZI, ELIZABETH	80	14	1	95	6.7	0.2	31.9
KORMAN, MICHAEL	25	28	2	55	6.7	0.8	22.1
HOHOTS, VIKTOR	403	95	6	504	5.9	2.2	12.5
Wang, Yaqian	9	23	1	33	4.2	0.1	21.1

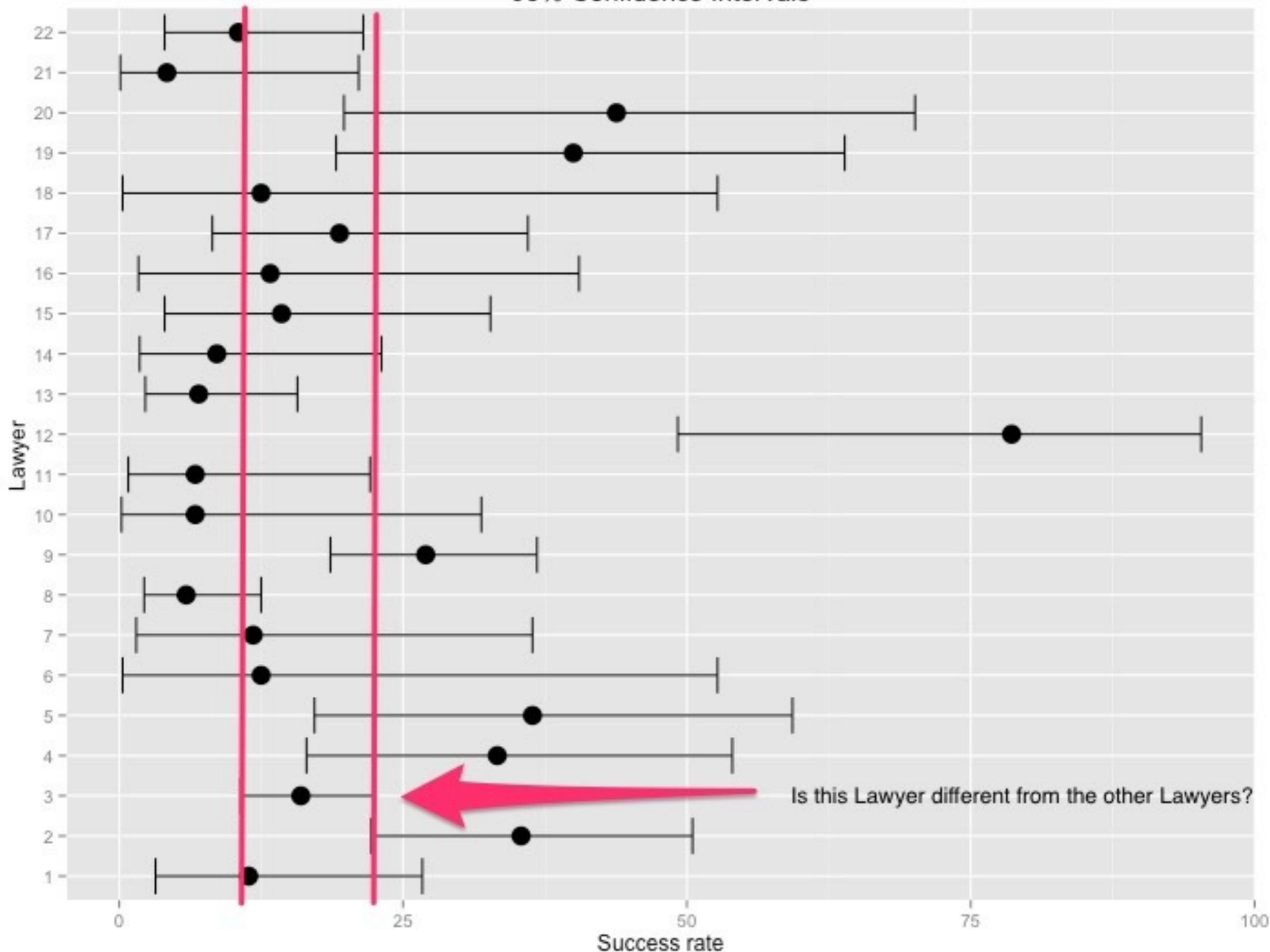
Table 2: 2008-2012 recognition rates for high volume counsel (25+ decisions) with 95% Confidence

Intervals - ranked by recognition rate

95% Confidence Intervals



95% Confidence Intervals



WHAT DO DATA VISUALIZATIONS SHOW?

- Patterns
- Relationships

STATE OF THE WORLD

Consider three estimates about the state of the world:

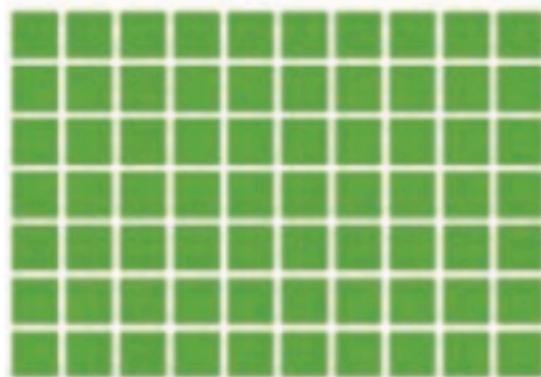
- A. Life expectancy at birth is 70 years
- B. The literacy rate of youth females ages 15 to 24 is 87 percent,
- C. The gross domestic product is approximately \$70 trillion.

Should you visualize this data?

Random numbers about the world

Life expectancy

70 years



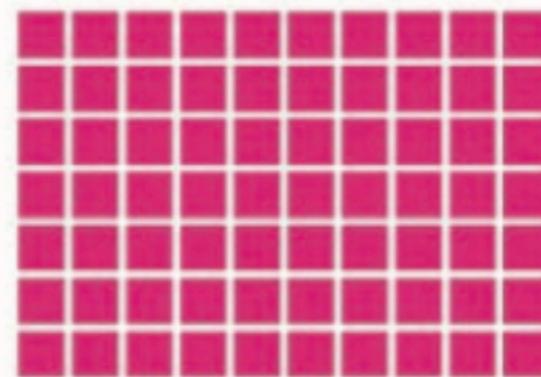
Literacy rate of youth females

87%



Gross domestic product

\$70 trillion



Random numbers about the world

LIFE EXPECTANCY

70 years

LITERACY RATE OF YOUTH FEMALES

87%

GROSS DOMESTIC PRODUCT

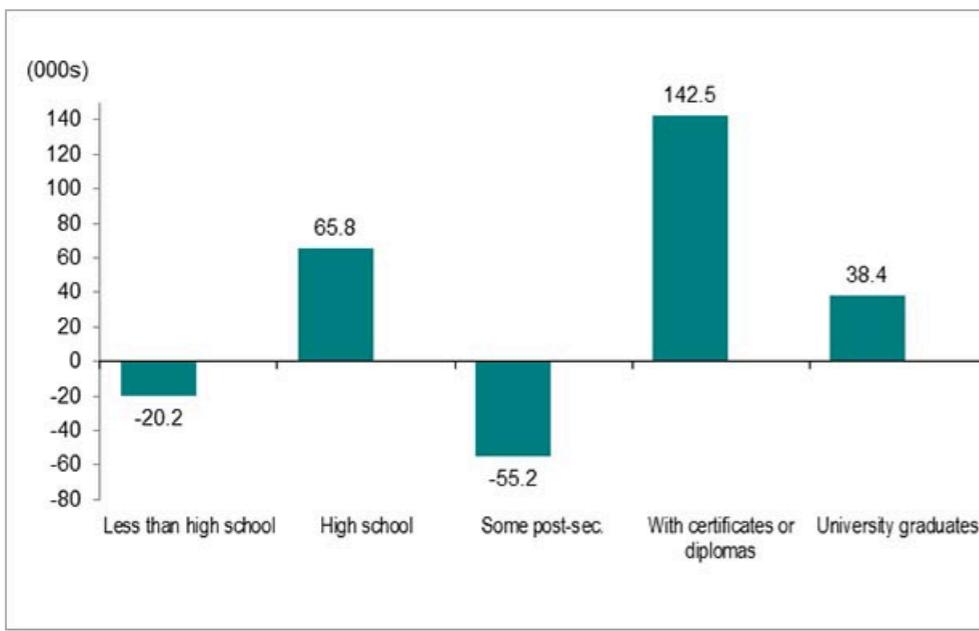
\$70 trillion

- Changes over time.
- Data can be split in different ways to reveal different patterns.

PATTERNS

Employment increase and decrease by education level

Chart 4 shows Ontario employment change by highest level of education attained, aged 25 and older, January 2018 to January 2019.

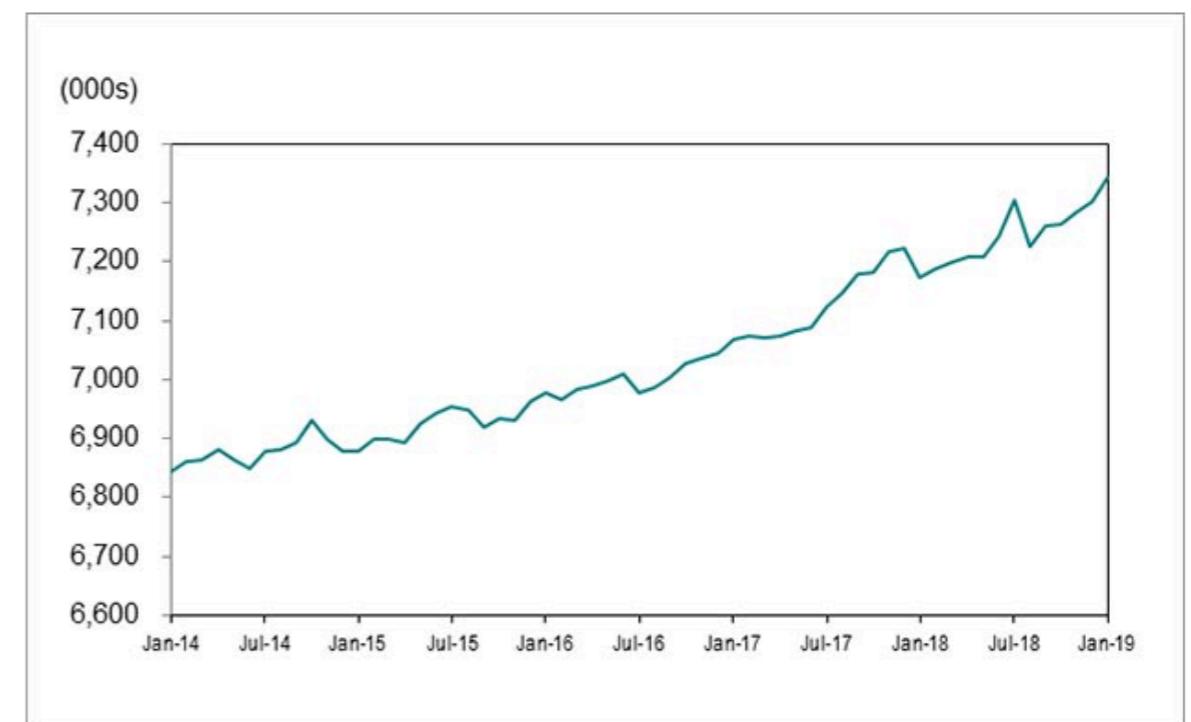


Source: Statistics Canada, Labour Force Survey, Table 14-10-0019-01, unadjusted data

Employment increased in January

Employment in Ontario increased in January (41,400), after rising by 16,100 jobs in December. January's job gain was the largest increase since July 2018.

Chart 1 shows employment in Ontario from January 2014 to January 2019.



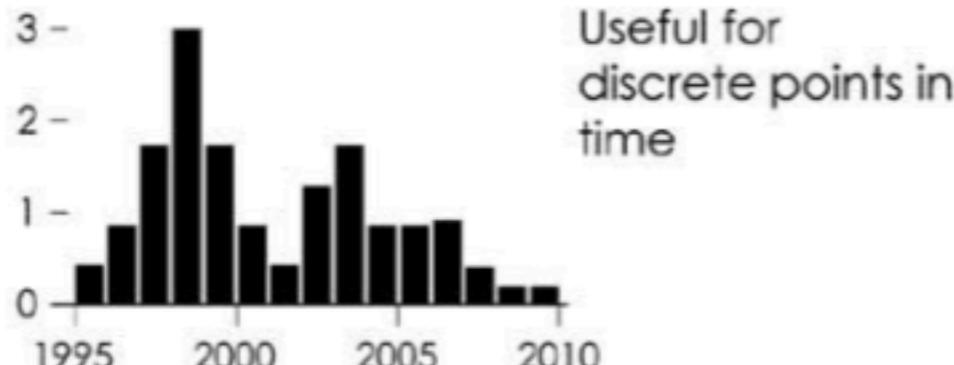
Source: Statistics Canada, Labour Force Survey, Table 14-10-0019-01, (seasonally adjusted data).

<https://www.ontario.ca/page/labour-market-report-january-2019>

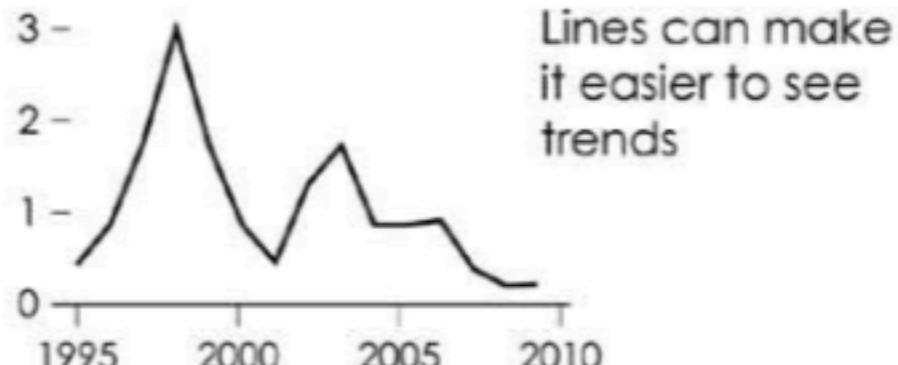
Time series

There are a variety of ways to see patterns over time, using cues such as length, direction, and position.

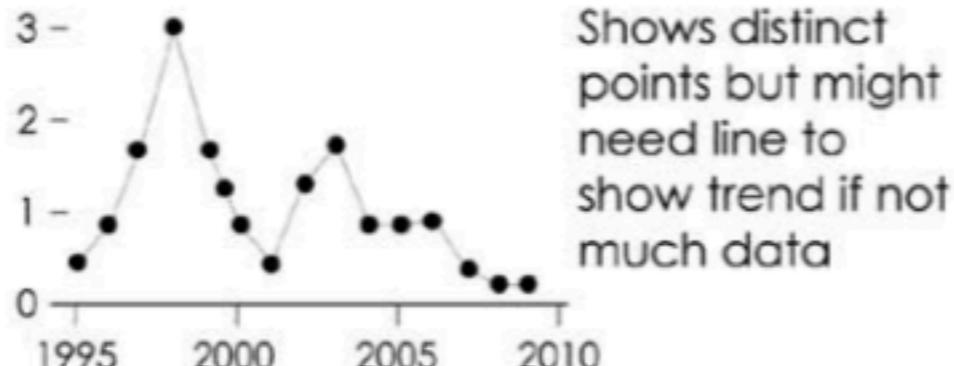
Bar graph



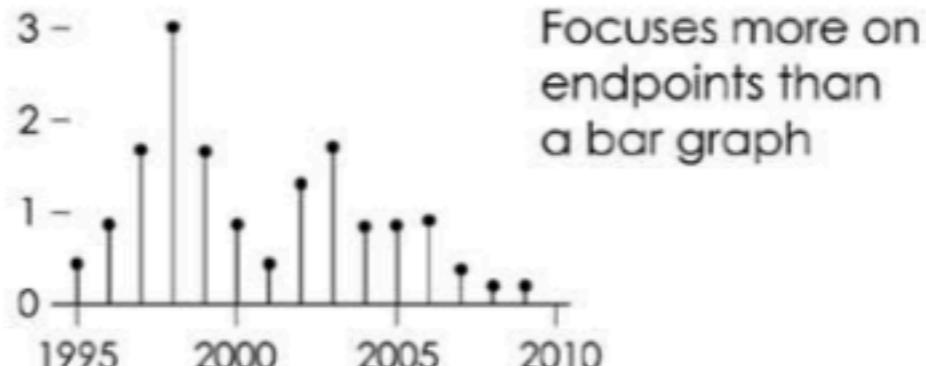
Line chart



Dot plot



Dot-bar graph

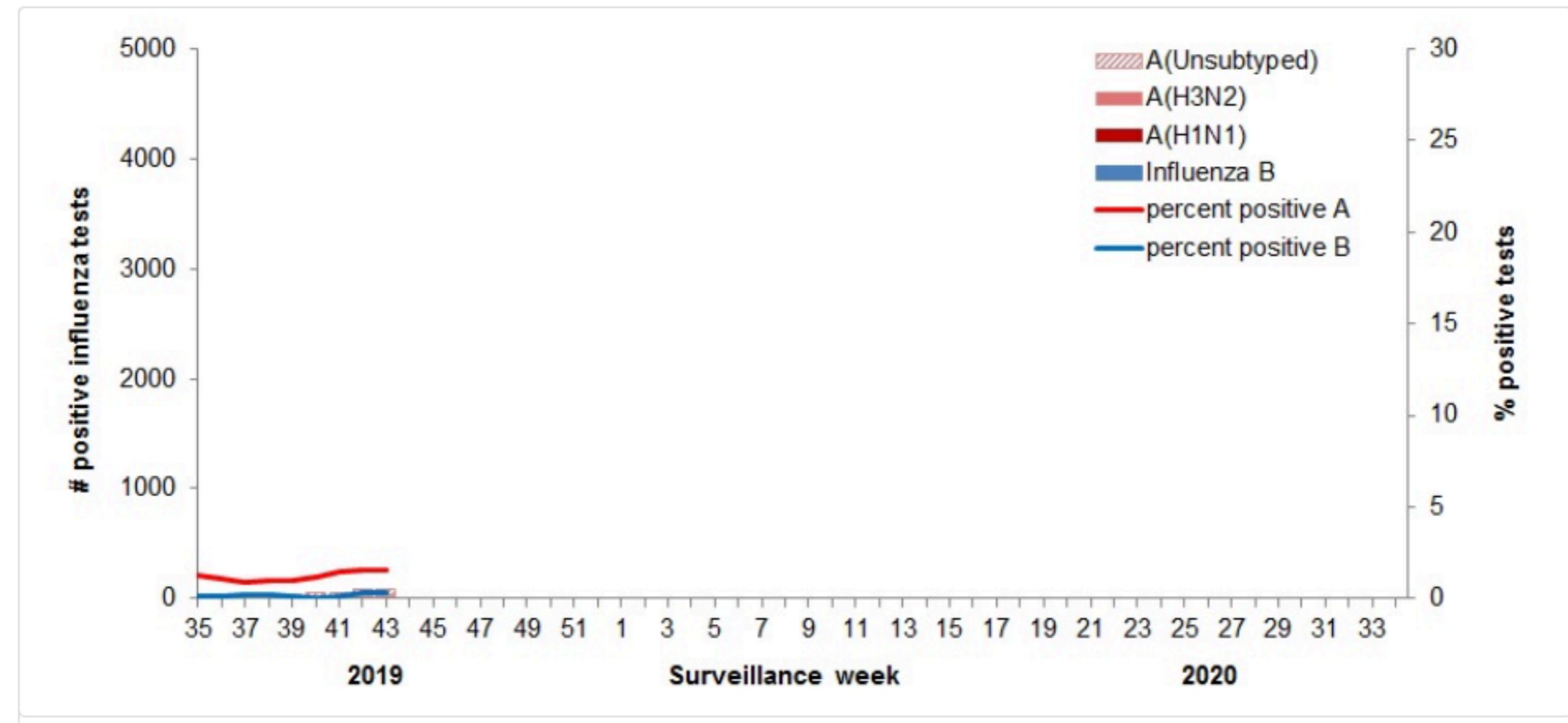


ANSWER THESE QUESTIONS BEFORE PRESENTING A VISUALIZATION ...

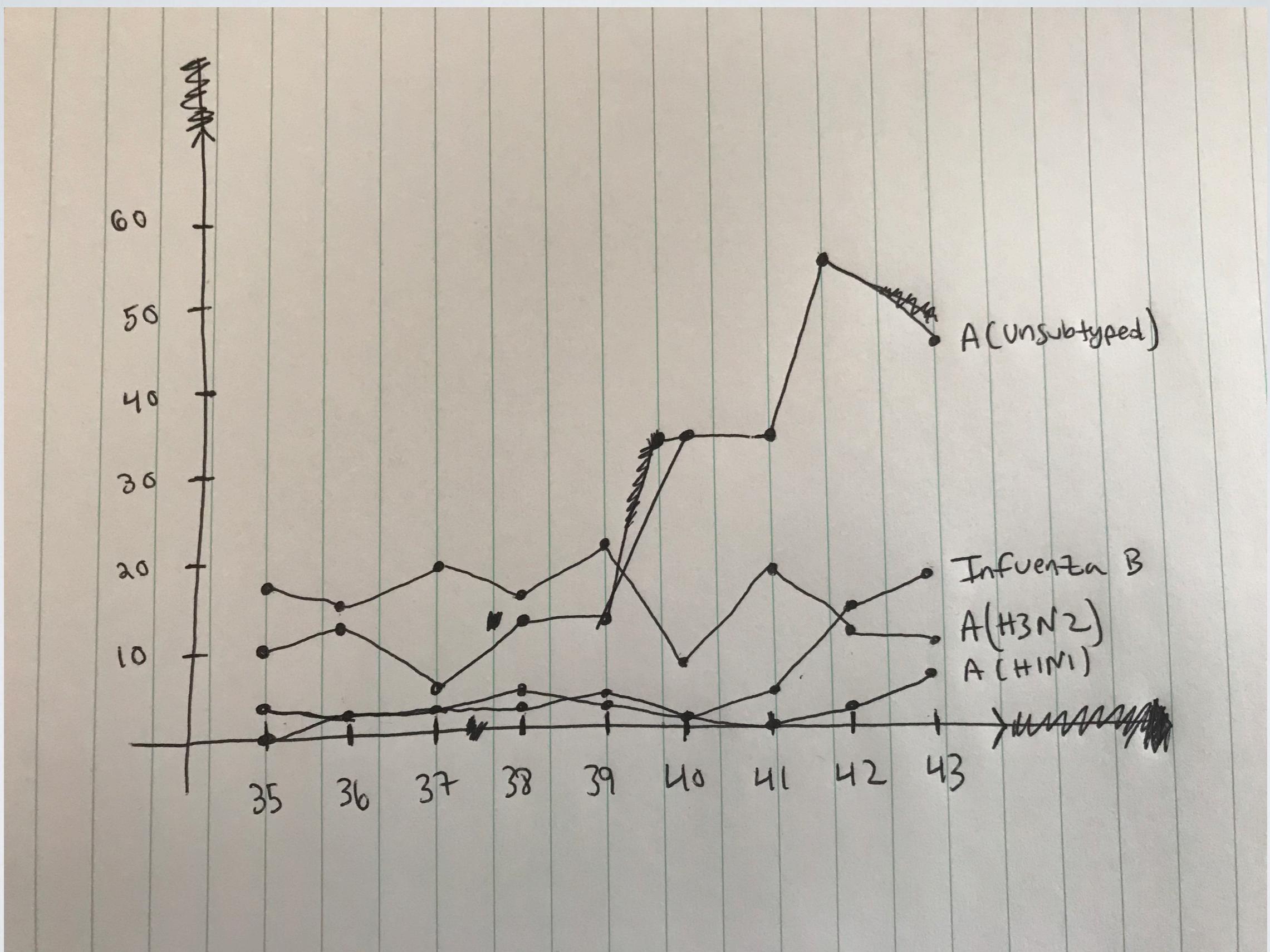
- What is your message?
- Who is your audience?
- What does your audience need to know?

Figure 2 - Number of positive influenza tests and percentage of tests positive, by type, subtype and report week, Canada, week 2019-43

Number of Laboratories Reporting in Week 43: 33 out of 34



<https://www.canada.ca/en/public-health/services/publications/diseases-conditions/fluwatch/2019-2020/week-43-october-20-26-2019.html>

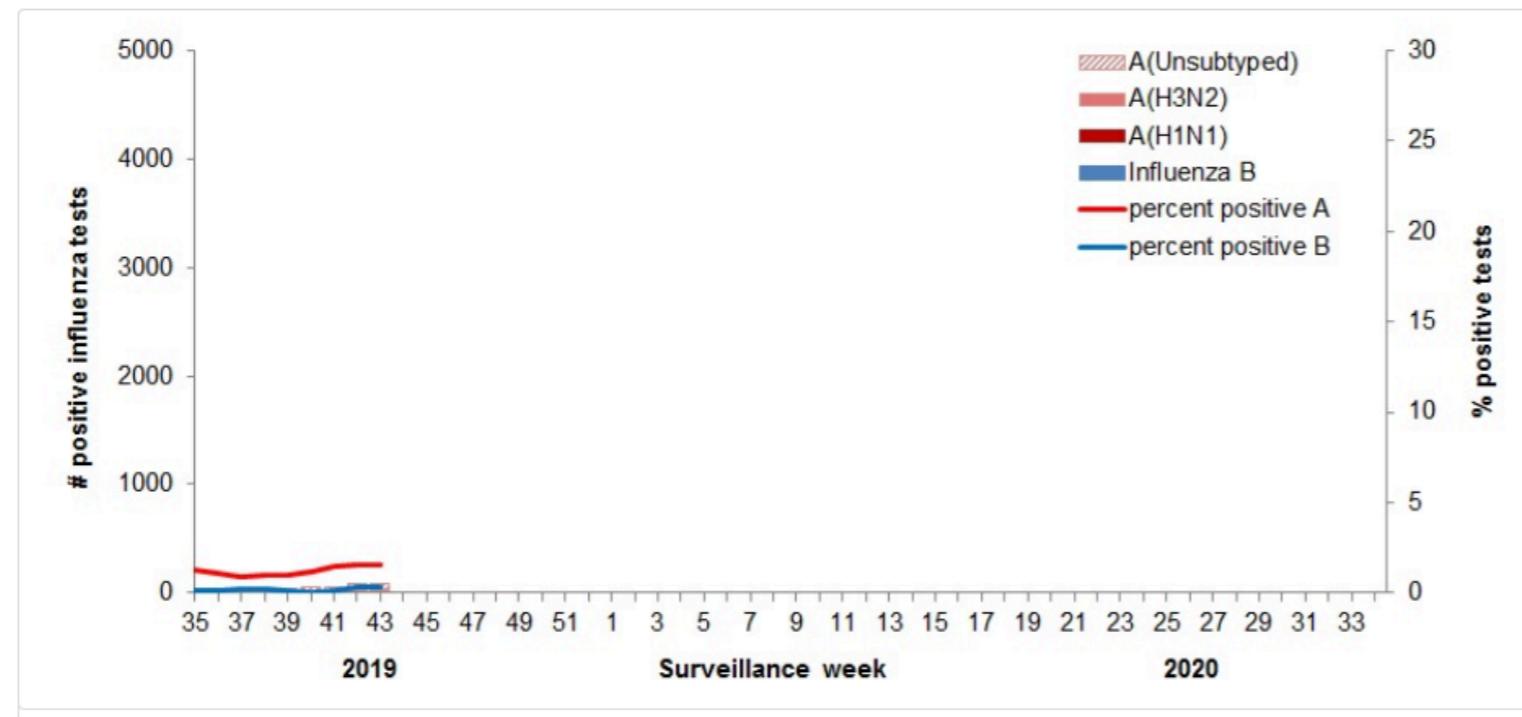


▼ Figure 2 - Text equivalent

Surveillance Week	A(Unsubtyped)	A(H3N2)	A(H1N1)pdm09	Influenza B	Percent Positive A	Percent Positive B
35	10	16	0	2	1.3	0.1
36	11	13	2	2	1.1	0.1
37	5	17	2	5	0.9	0.2
38	11	15	3	6	1.0	0.2
39	11	21	2	3	1.0	0.1
40	34	9	1	2	1.2	0.1
41	34	18	0	5	1.4	0.1
42	54	12	1	14	1.6	0.3
43	45	12	6	17	1.6	0.3

Figure 2 - Number of positive influenza tests and percentage of tests positive, by type, subtype and report week, Canada, week 2019-43

Number of Laboratories Reporting in Week 43: 33 out of 34



Lab Confirmed Influenza in Week 43

Source: Government of Canada FluWatch

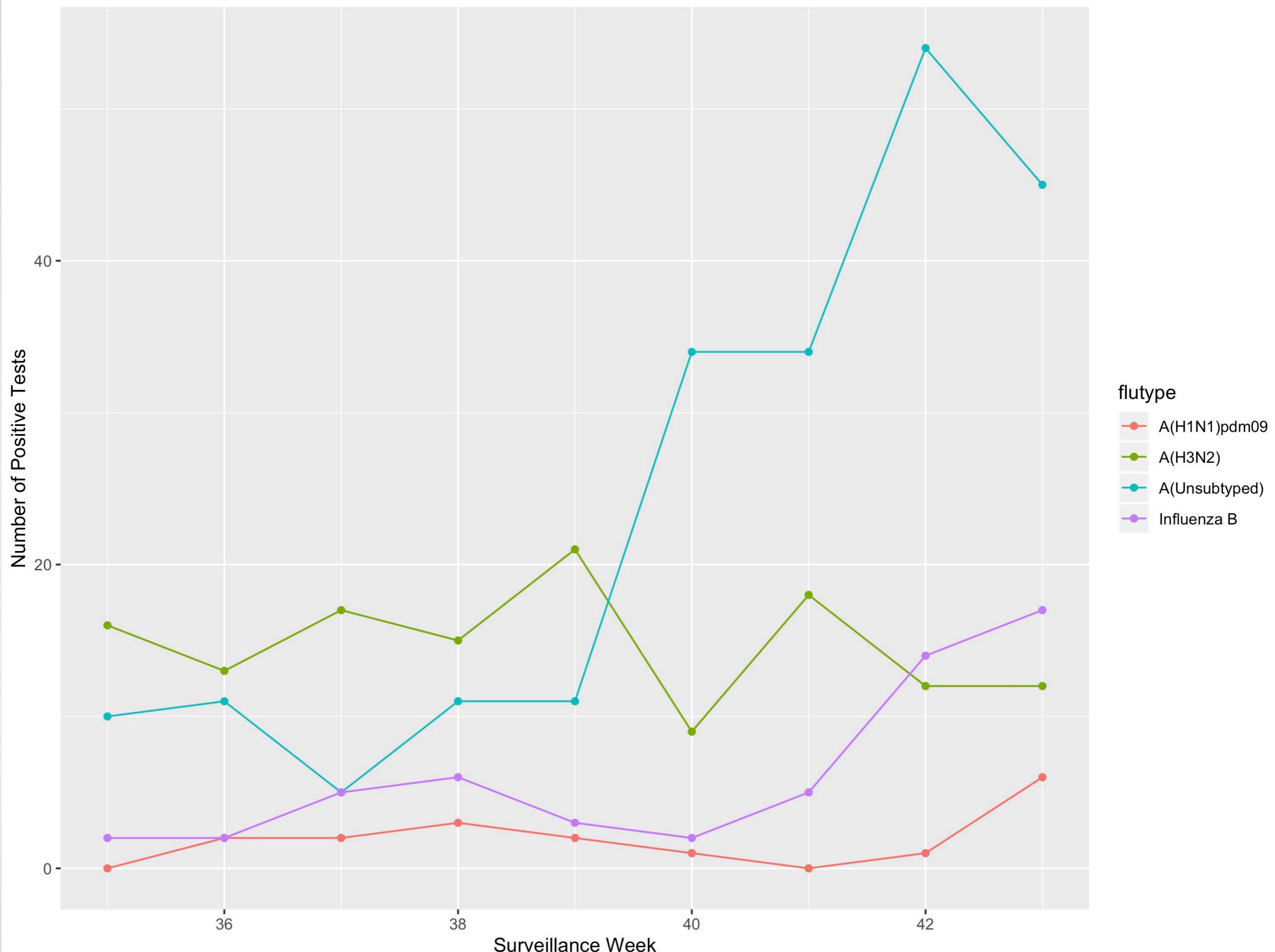
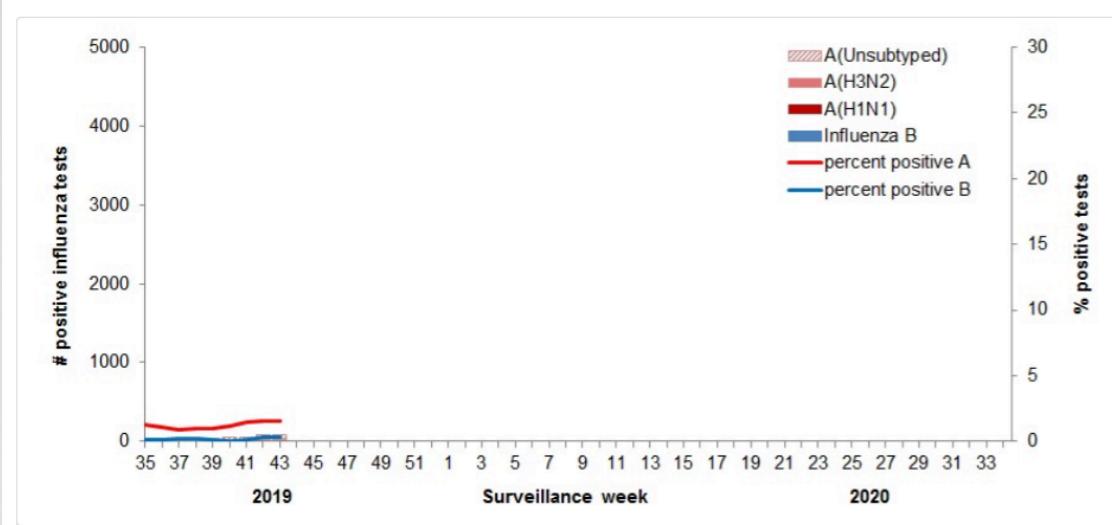


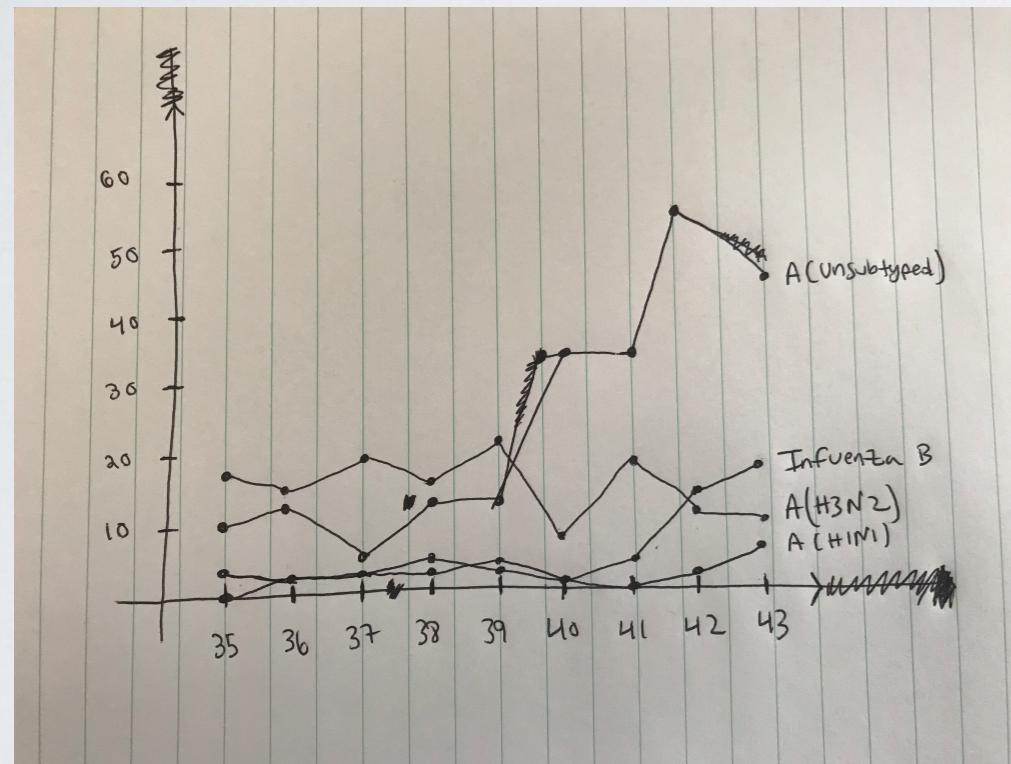
Figure 2 - Number of positive influenza tests and percentage of tests positive, by type, subtype and report week, Canada, week 2019-43

Number of Laboratories Reporting in Week 43: 33 out of 34



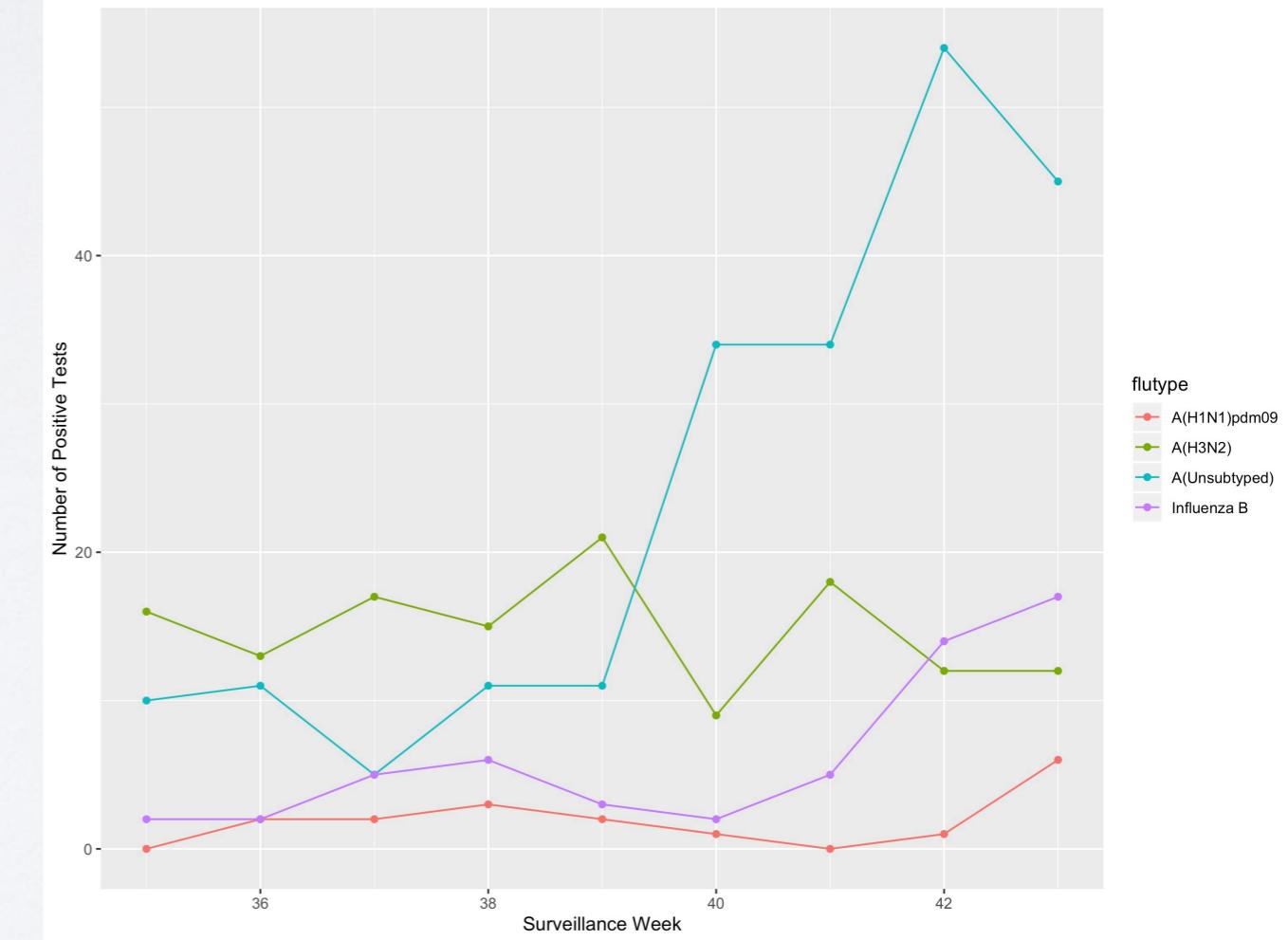
▼ Figure 2 - Text equivalent

Surveillance Week	A(Unsubtyped)	A(H3N2)	A(H1N1)pdm09	Influenza B	Percent Positive A	Percent Positive B
35	10	16	0	2	1.3	0.1
36	11	13	2	2	1.1	0.1
37	5	17	2	5	0.9	0.2
38	11	15	3	6	1.0	0.2
39	11	21	2	3	1.0	0.1
40	34	9	1	2	1.2	0.1
41	34	18	0	5	1.4	0.1
42	54	12	1	14	1.6	0.3
43	45	12	6	17	1.6	0.3



Lab Confirmed Influenza in Week 43

Source: Government of Canada FluWatch



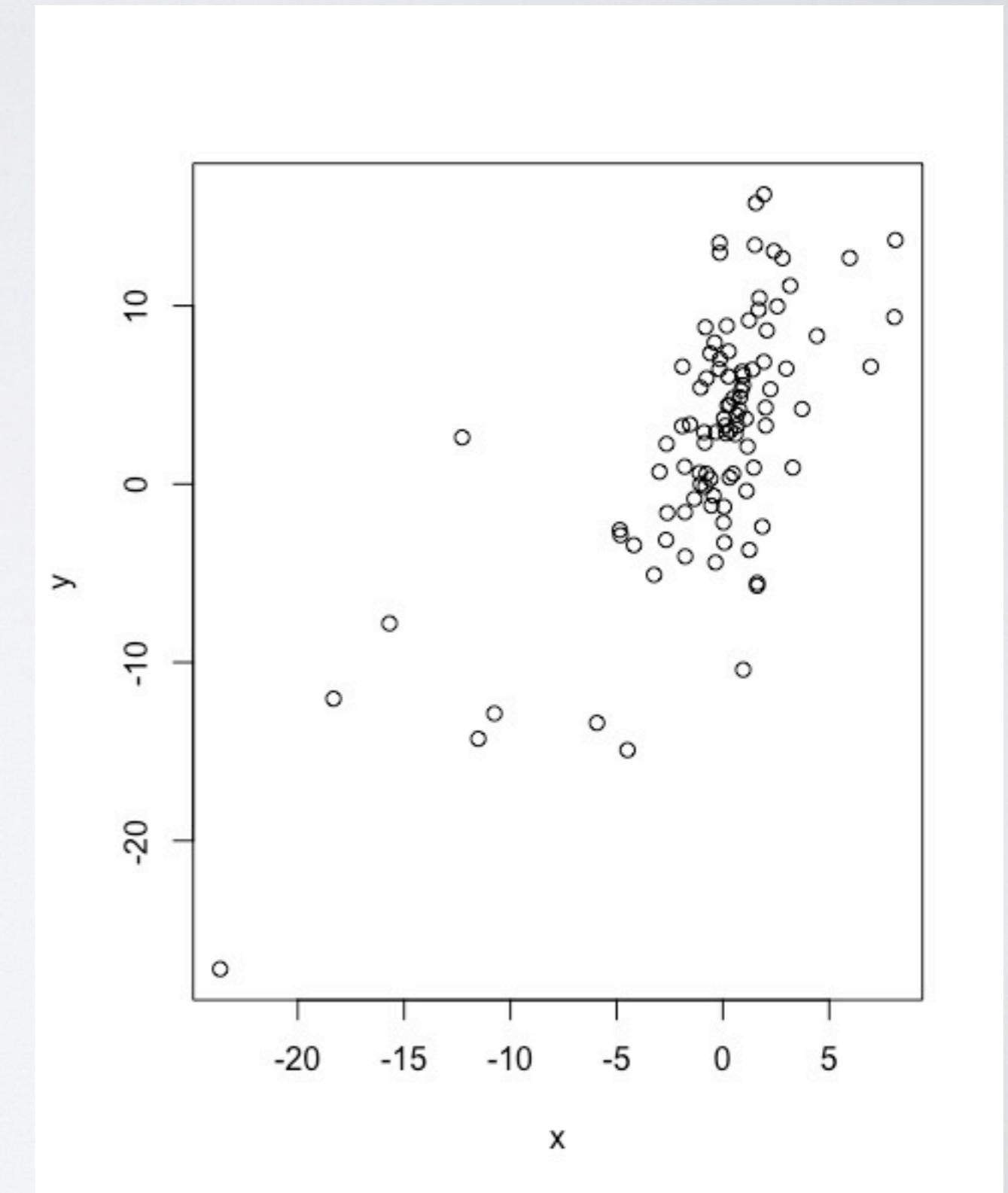
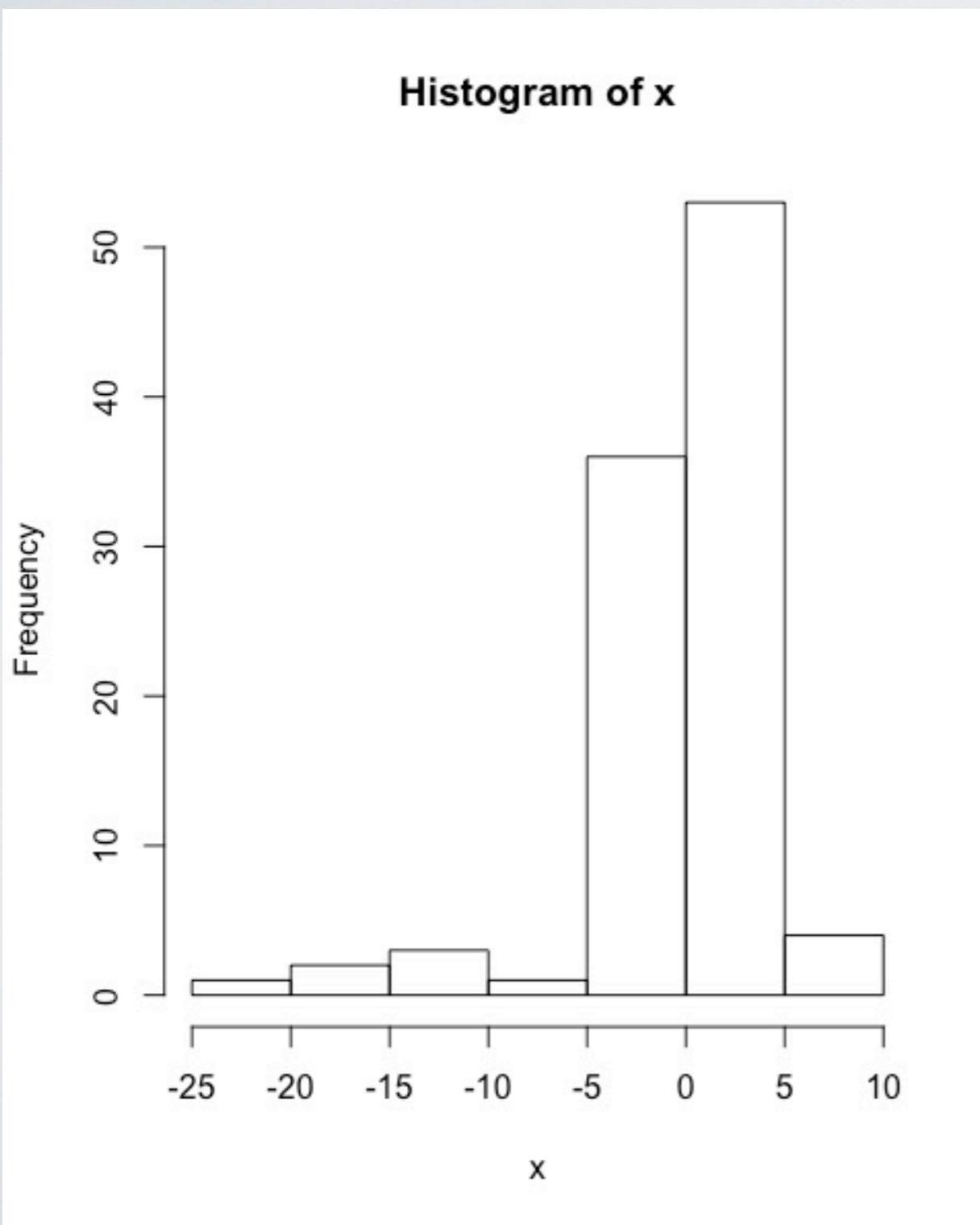
PRESENTING DATA TO PEOPLE

- Me, myself, and I
- A specific audience
- A wider audience

PRESENTING DATA TO PEOPLE

- How much control does the audience have over the presentation?
- How much detail can they get?

ME, MYSELF, AND I

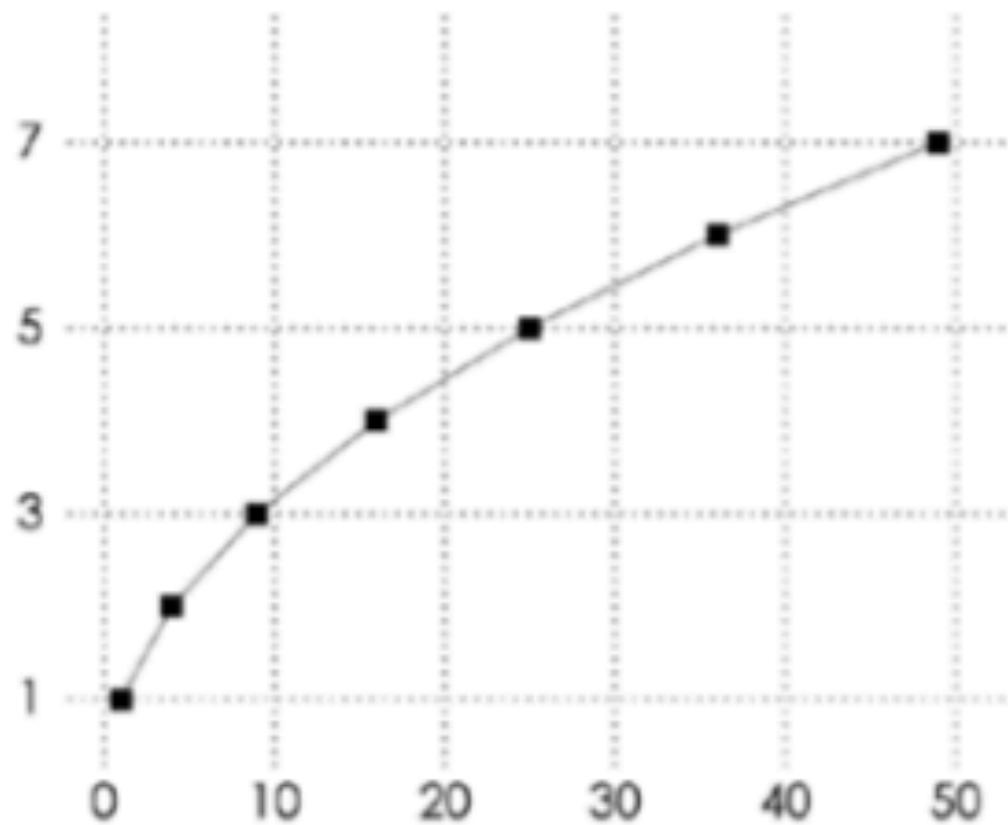


A SPECIFIC AUDIENCE

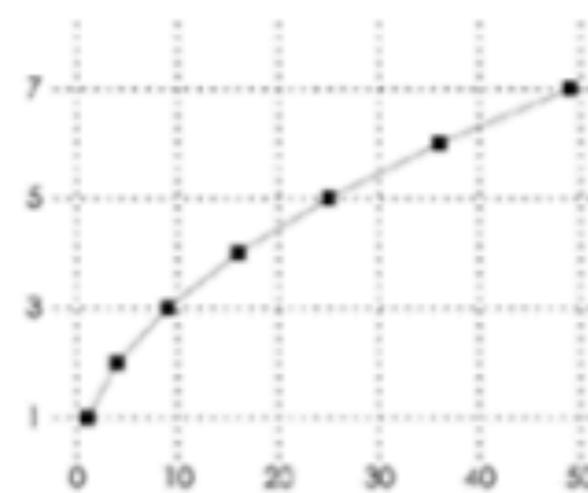
- Your audience should be able to decode your encodings so that they can understand the data.
- If your audience is already familiar with the background behind your data or has perhaps even worked with it, the barriers are lower, but still exist.
- Consider how your audience will examine your work.

VISUALIZATION IN A PRESENTATION

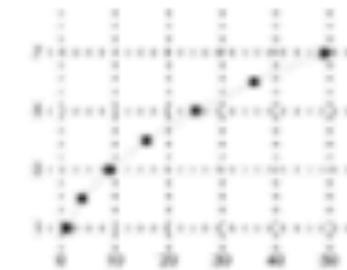
You can see this okay.



This too, if you squint.



Um, what?



DESIGNING FOR A WIDER AUDIENCE

- As your audience grows so do the challenges, such as the range of data literacy, and familiarity with your data's context.
- Avoid jargon and be sure you explain complex concepts in a way so that people can relate.

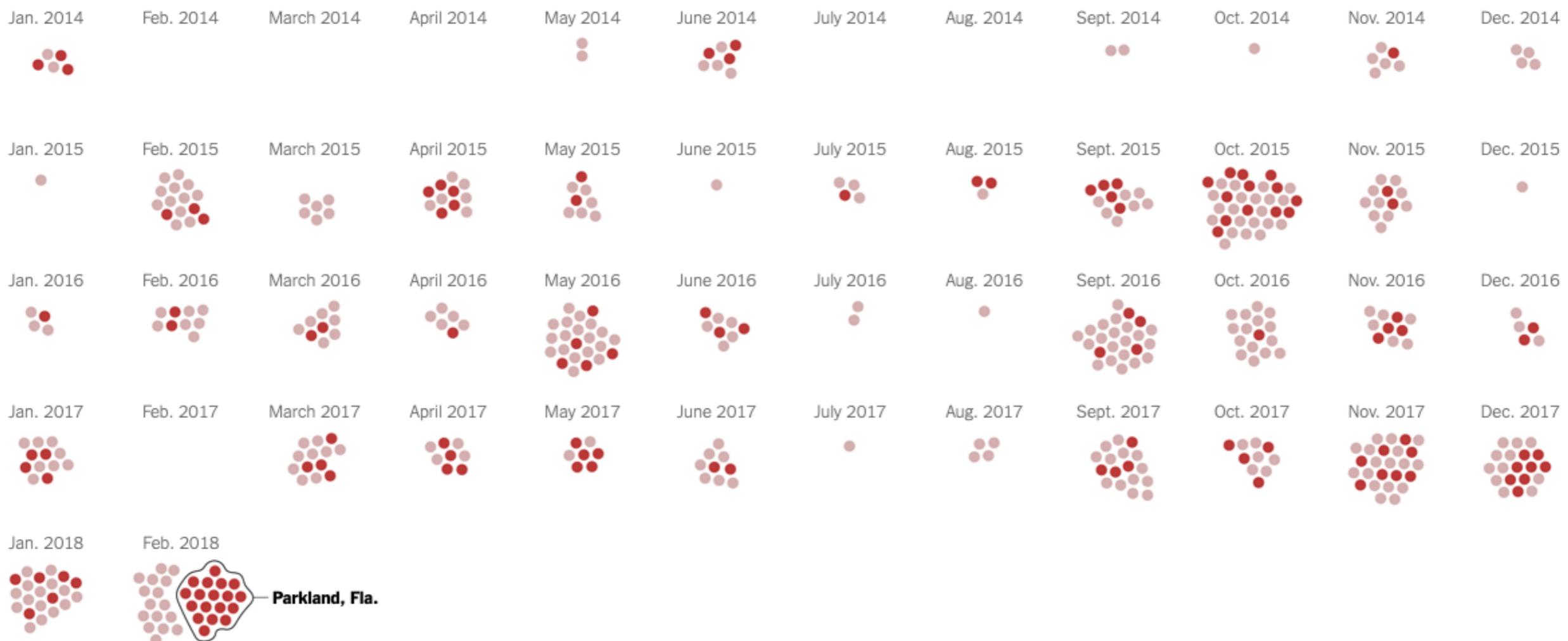
After Sandy Hook, More Than 400 People Have Been Shot in Over 200 School Shootings

By JUGAL K. PATEL FEB. 15, 2018

Gunshot Victims in School Shootings

● Killed ● Injured

Sandy Hook



Source: Gun Violence Archive

Note: Shootings in 2013 are not included because complete data was not available in that year. Months with blanks indicate no shootings archived.

THINGS TO CONSIDER

- Imagine you are a tourist in a new place.
- What do you want a tour guide to tell you?
- It's your job to point out the direction of interest, provide background, and make sure you don't confuse people.

DATA PROVENANCE

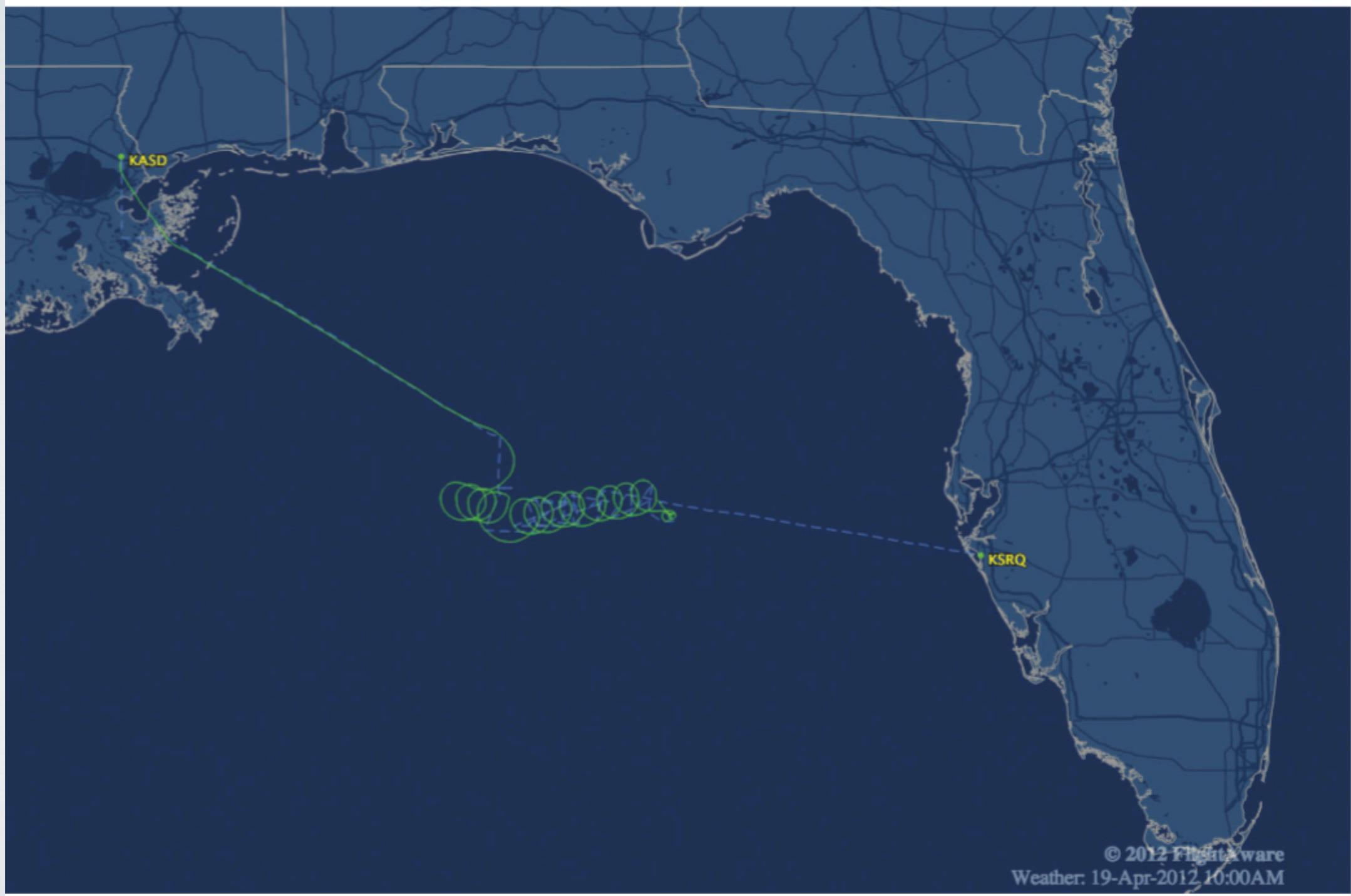
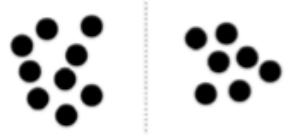
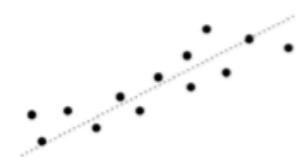


FIGURE 6-17 A flight from Slidell, Louisiana to Sarasota, Florida, according to FlightAware, <http://flightaware.com/live/flight/N48DL>

DATA NARRATIVE

- Ask a question about the data and then try to answer that through the visualization.
- How do you want your audience to read the data? How will your audience read your graph?

Possible questions Fill in the blanks	Statistical concepts	Possible visuals
What _____ is the best and worst?	Maximums and minimums	
How has _____ changed over time?	Temporal patterns	
What _____ stands out from the rest?	Outliers	
What makes _____ different from _____?	Clustering	
How are _____ and _____ related to each other?	Correlation	
What's the breakdown for _____?	Distributions	



Donald J. Trump 
@realDonaldTrump



214K 6:05 AM - Oct 1, 2019



110K people are talking about this



<https://www.cnn.com/2019/10/01/politics/donald-trump-map-2016-election/index.html>

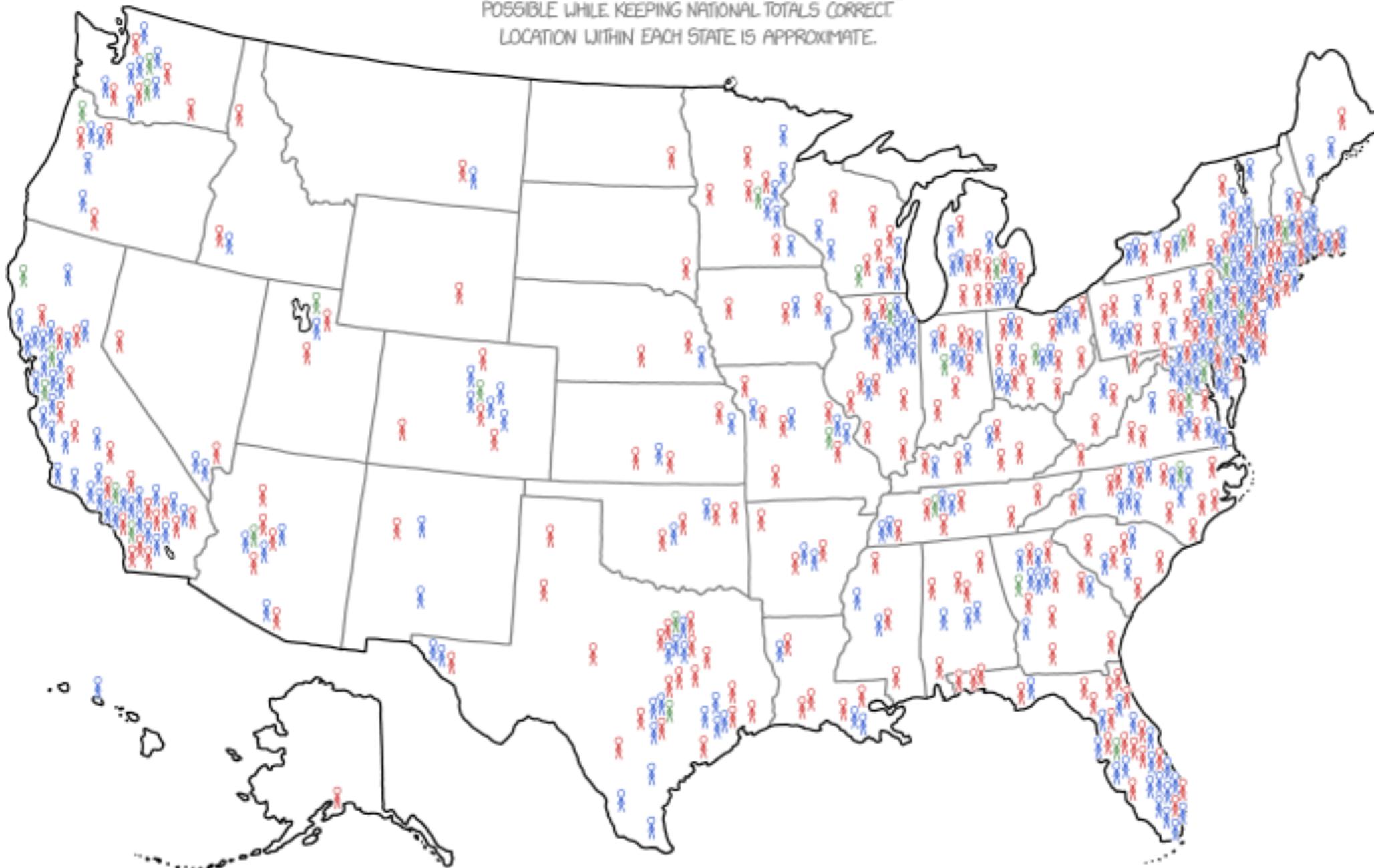
2016 ELECTION MAP

EACH FIGURE REPRESENTS 250,000 VOTES

TRUMP CLINTON OTHER

VOTES ARE DISTRIBUTED BY STATE AS ACCURATELY AS
POSSIBLE WHILE KEEPING NATIONAL TOTALS CORRECT.

LOCATION WITHIN EACH STATE IS APPROXIMATE.



United States presidential election, 2008



2004 ← November 4, 2008 → 2012

All 538 electoral votes of the Electoral College

270 electoral votes needed to win

Turnout

58.2%^[1] ▲ 1.5%



Nominee Barack Obama

Party Democratic

Home state Illinois

Running mate Joe Biden

Electoral vote 365

States carried 28 + DC + NE-02

Popular vote 69,498,516

Percentage 52.9%

Nominee John McCain

Party Republican

Home state Arizona

Running mate Sarah Palin

Electoral vote 173

States carried 22

Popular vote 59,948,323

Percentage 45.7%

United States presidential election, 2012



2008 ← November 6, 2012 → 2016

All 538 electoral votes of the Electoral College

270 electoral votes needed to win

Turnout

54.9%^[1] ▼ 3.3%



Nominee Barack Obama

Party Democratic

Home state Illinois

Running mate Joe Biden

Electoral vote 332

States carried 26 + DC

Popular vote 65,915,795

Percentage 51.1%

Nominee Mitt Romney

Party Republican

Home state Massachusetts

Running mate Paul Ryan

Electoral vote 206

States carried 24

Popular vote 60,933,504

Percentage 47.2%

United States presidential election, 2016



2012 ← November 8, 2016 → 2020

538 members of the Electoral College

270 electoral votes needed to win



Nominee Hillary Clinton

Party Democratic

Home state New York

Running mate Tim Kaine

Projected electoral vote 232^{[1][2][3]}

States carried 20 + DC

Popular vote 59,861,516^[4]

Percentage 47.7%

Nominee Donald Trump

Party Republican

Home state New York

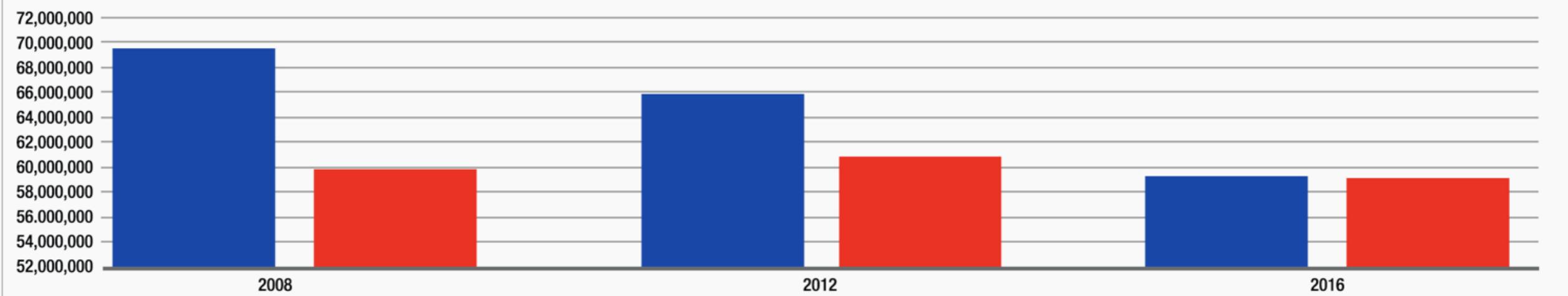
Running mate Mike Pence

Electoral vote 306^{[1][2][3]}

States carried 30 + ME-02

Popular vote 59,639,462^[4]

Percentage 47.5%

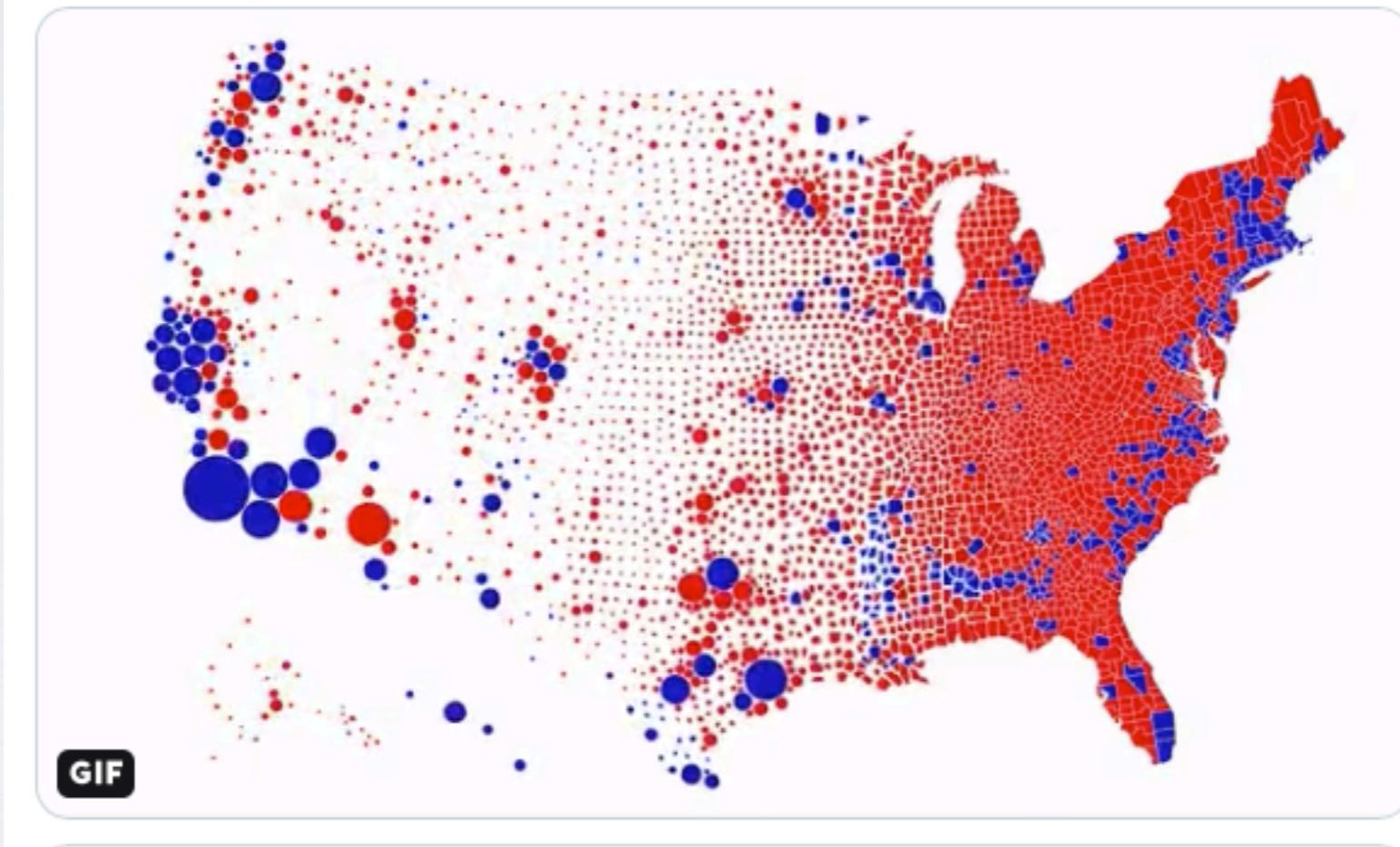


source: https://en.wikipedia.org/wiki/United_States_presidential_election,_2008
https://en.wikipedia.org/wiki/United_States_presidential_election,_2012
https://en.wikipedia.org/wiki/United_States_presidential_election,_2016



Karim Douieb
@karim_douieb

Challenge accepted! Here is a transition between surface area of US counties and their associated population. This arguably provides a much more accurate reading of the situation. [@observablehq notebook](#):
observablehq.com/@karimdouieb/t... #HowChartsLie
#DataViz #d3js



https://twitter.com/karim_douieb/status/1181695687005745153