GENERAL ASSEMBLY

CREATING DATA NARRATIVES WITH DATA VISUALIZATION

Joseph Nelson

AGENDA

- Introduction
- Why visualize?
- Data visualization best practices
- Tools of the trade

INTRO(S)

(DON'T TAKE NOTES ON THIS PART)

- Co-Founder, Represently
- **From:** Des Moines, Iowa
- **Influences:** a16z, Andrew Ng
- **Likes:** Hockey, bad data science puns, biking, data viz, new tech (especially for good)

FLEISHMANHILLARD

USDA

GA

GW

WHY

FROM 30,000 FEET

STATE	TOTAL VOTES	Votes	%	EV	Votes	%	EV	Votes	%	EV
Alabama	2,123,372	729,547	34.4%		1,318,255	62.1%	9	44,467	2.1%	
Alaska	318,608	116,454	36.6%		163,387	51.3%	3	18,725	5.9%	
Arizona	2,573,165	1,161,167	45.1%		1,252,401	48.7%	11	106,327	4.1%	
Arkansas	1,130,635	380,494	33.7%		684,872	60.6%	6	29,829	2.6%	
California	14,181,595	8,753,788	61.7%	55	4,483,810	31.6%		478,500	3.4%	
Colorado	2,780,220	1,338,870	48.2%	9	1,202,484	43.3%		144,121	5.2%	
Connecticut	1,644,920	897,572	54.6%	7	673,215	40.9%		48,676	3.0%	
Delaware	441,590	235,603	53.4%	3	185,127	41.9%		14,757	3.3%	
Dist. of Col.	311,268	282,830	90.9%	3	12,723	4.1%		4,906	1.6%	
Florida	9,420,039	4,504,975	47.8%		4,617,886	49.0%	29	207,043	2.2%	
Georgia	4,092,373	1,877,963	45.9%		2,089,104	51.0%	16	125,306	3.1%	
Hawaii	428,937	266,891	62.2%	3*	128,847	30.0%		15,954	3.7%	
Idaho	690,255	189,765	27.5%		409,055	59.3%	4	28,331	4.1%	
Illinois	5,536,424	3,090,729	55.8%	20	2,146,015	38.8%		209,596	3.8%	
Indiana	2,734,958	1,033,126	37.8%		1,557,286	56.9%	11	133,993	4.9%	
Iowa	1,566,031	653,669	41.7%		800,983	51.1%	6	59,186	3.8%	
Kansas	1,184,402	427,005	36.1%		671,018	56.7%	6	55,406	4.7%	
Kentucky	1,924,149	628,854	32.7%		1,202,971	62.5%	8	53,752	2.8%	
Louisiana	2,029,032	780,154	38.4%		1,178,638	58.1%	8	37,978	1.9%	
Maine	747,927	357,735	47.8%	3	335,593	44.9%	1	38,105	5.1%	
Maryland	2,781,446	1,677,928	60.3%	10	943,169	33.9%		79,605	2.9%	
Massachusetts	3,325,046	1,995,196	60.0%	11	1,090,893	32.8%		138,018	4.2%	
Michigan	4,799,284	2,268,839	47.3%		2,279,543	47.5%	16	172,136	3.6%	
Minnesota	2,944,813	1,367,716	46.4%	10	1,322,951	44.9%		112,972	3.8%	
Mississippi	1,209,357	485,131	40.1%		700,714	57.9%	6	14,435	1.2%	
Missouri	2,808,605	1,071,068	38.1%		1,594,511	56.8%	10	97,359	3.5%	
Montana	494,526	177,709	35.9%		279,240	56.5%	3	28,037	5.7%	
Nebraska	844,227	284,494	33.7%		495,961	58.7%	2	38,946	4.6%	
CD-1	*282,338*	*100,126*	*35.5%*		*158,626*	*56.2%*	1	*14,031*	*5.0%*	
CD-2	*291,680*	*131,030*	*44.9%*		*137,564*	*47.2%*	1	*13,245*	*4.5%*	
CD-3	*270,109*	*53,290*	*19.7%*		*199,657*	*73.9%*	1	*11,657*	*4.3%*	
Nevada	1,125,385	539,260	47.9%	6	512,058	45.5%		37,384	3.3%	
New Hampshire	744,296	348,526	46.8%	4	345,790	46.5%		30,777	4.1%	
New Jersey	3,874,046	2,148,278	55.5%	14	1,601,933	41.4%		72,477	1.9%	

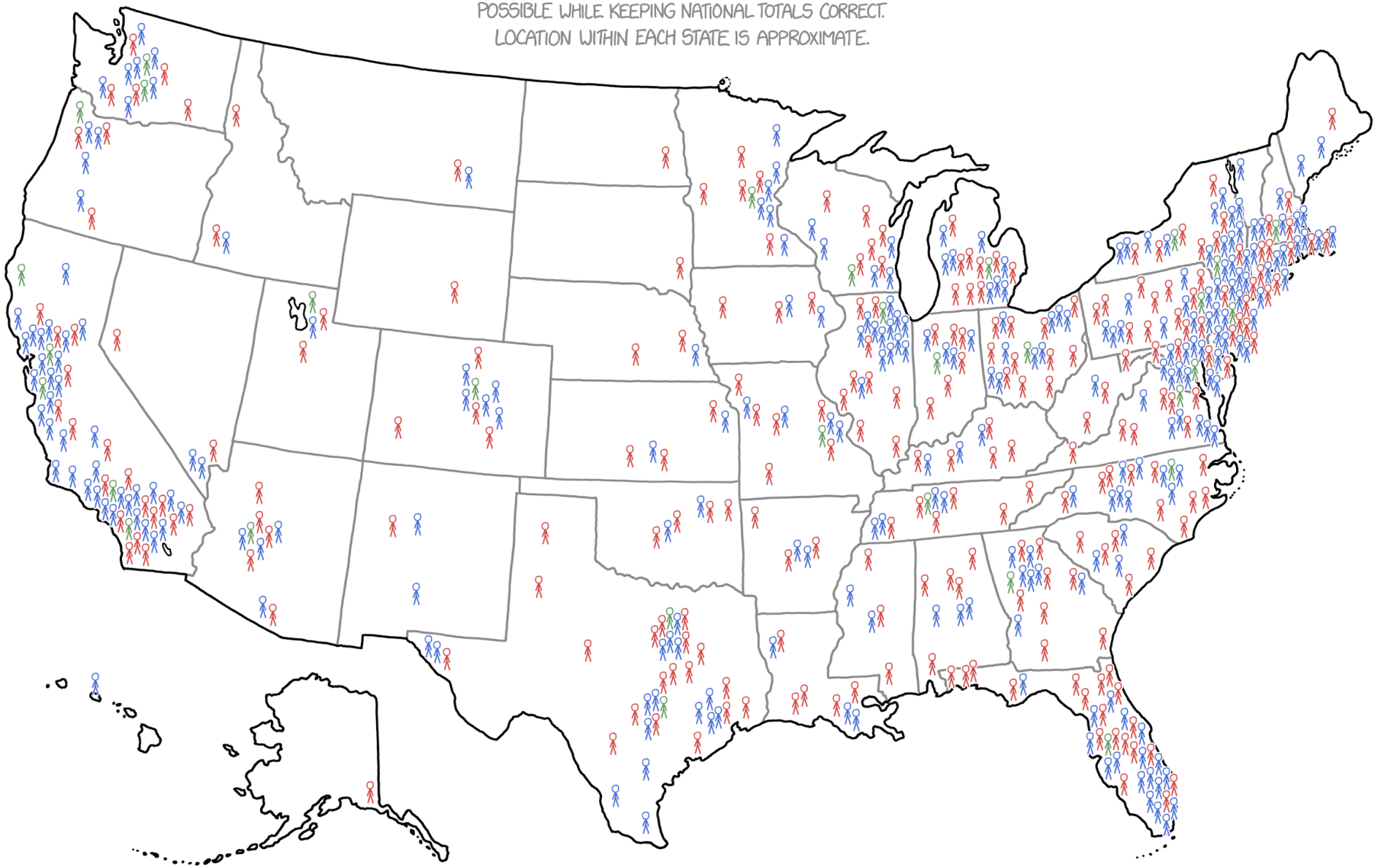
2016 ELECTION MAP

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.



Why visualize data?

1. **Discover** previously unexplored relationships.

2. **Communicate** with evidence.

HOW

Data visualization starts with...

Data visualization starts with...

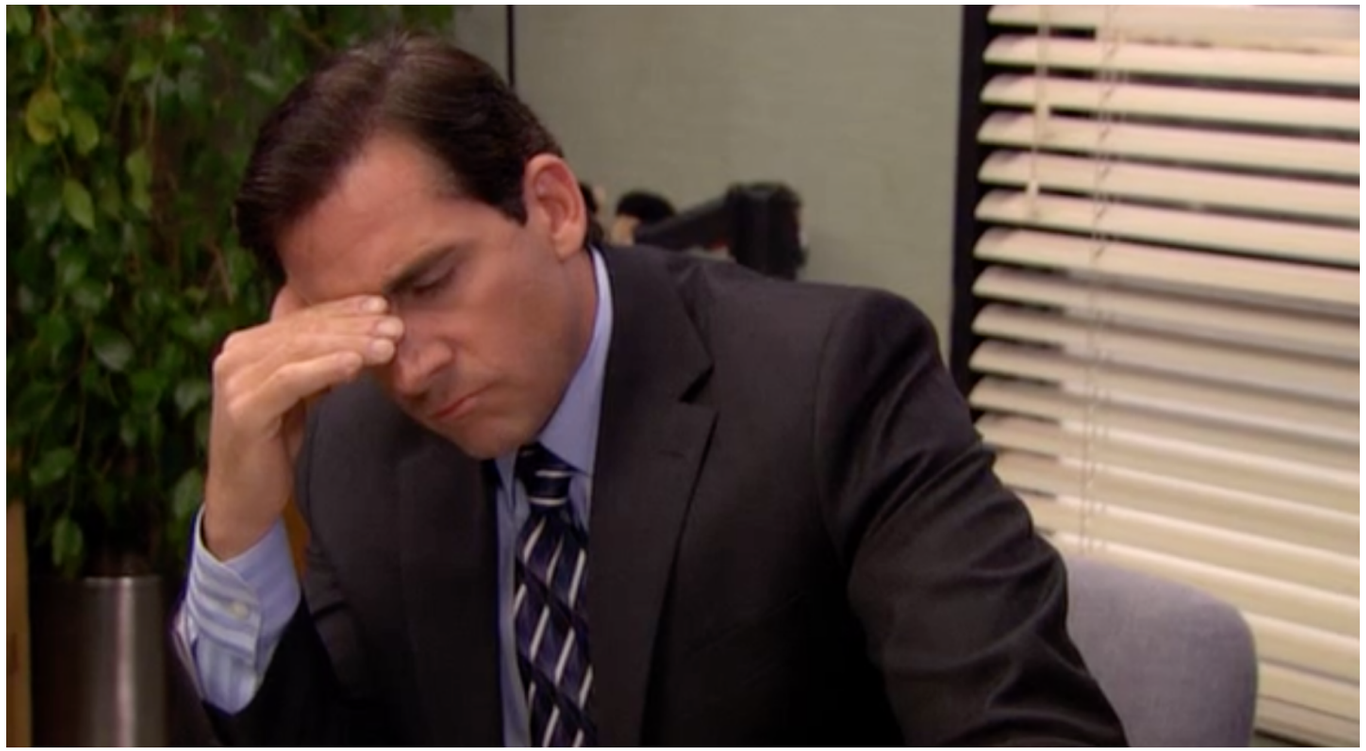
Data visualization starts with...

Narrative.

Data visualization starts with a narrative

- You must have a hypothesis you want to prove, or disprove.
- You need to start with why.
- You need to understand your audience.
- Visualizations aid your established narrative.

Start with why, tailored to your audience



- https://vimeo.com/27060669

Let’s look at examples.

- 1. How Americans Spend their Days (Flowing Data)
- 2. Over the Decades, How States Have Shifted (New York Times)
- 3. The Data Behind a Season without Snow Days (Medium, self)
- 4. Why Underdogs Do Better in Hockey Than Basketball (Vox)
- 5. Messi Is Impossible (FiveThirtyEight)

Data journalism vs data visualization

- In data journalism, data visuals can be **the story itself**.
- (How States Have Shifted)

- Data visualization, however, **aids any existing story**.
- (The Data Behind a Season without Snow Days)

If you like data journalism… (But we're going to focus on 'everyday' visuals)

Criteria for good visualization

What were the commons themes that created strong visuals?

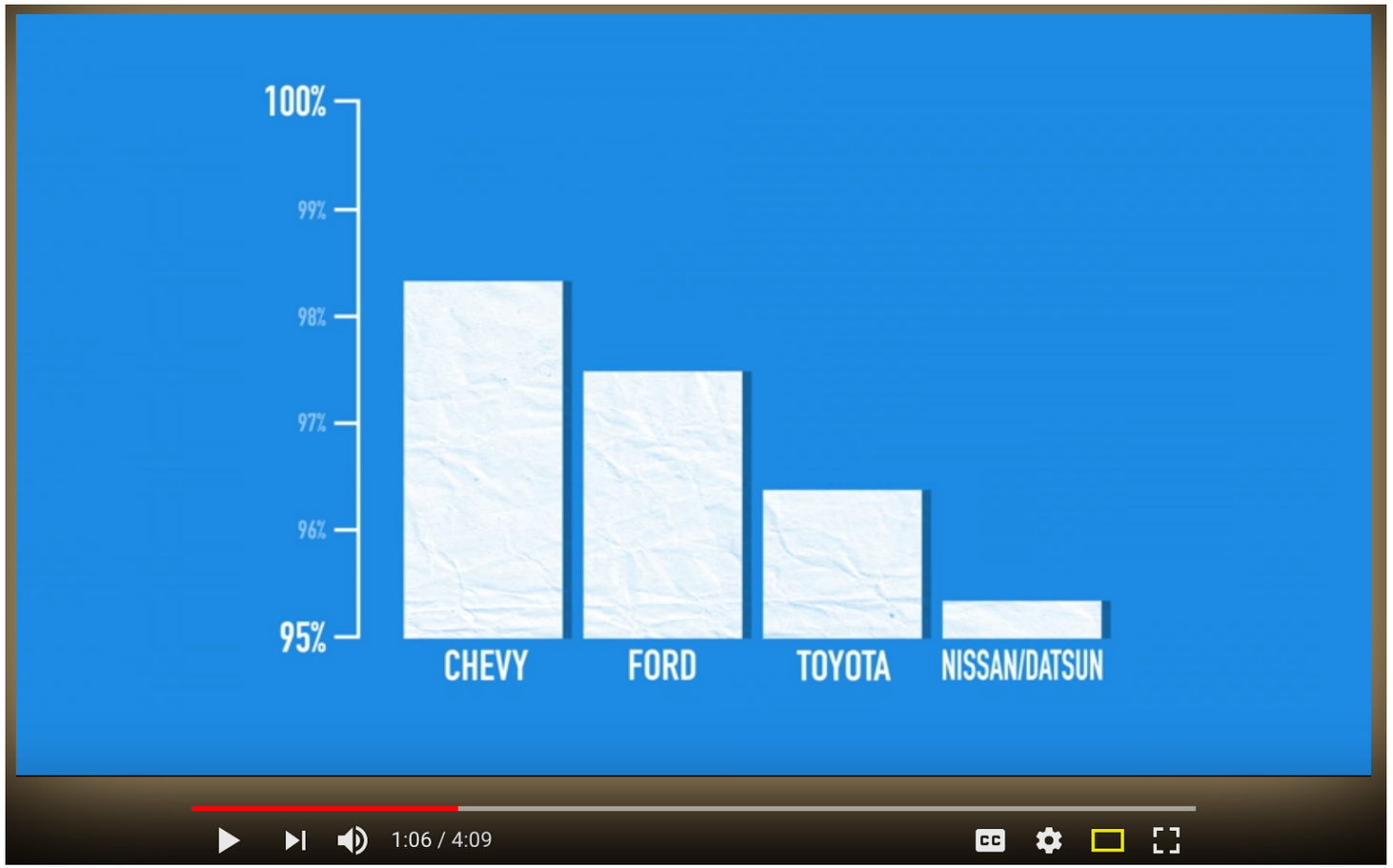
Criteria for good visualization

1. **Intuitive**. It passes the “10 second” test. We know our audience.

2. **Simplified.** Colors and relationships are clear.

3. **Clearly labeled.** The title is the takeaway. Our axes are labeled.

Yes, axes matter.



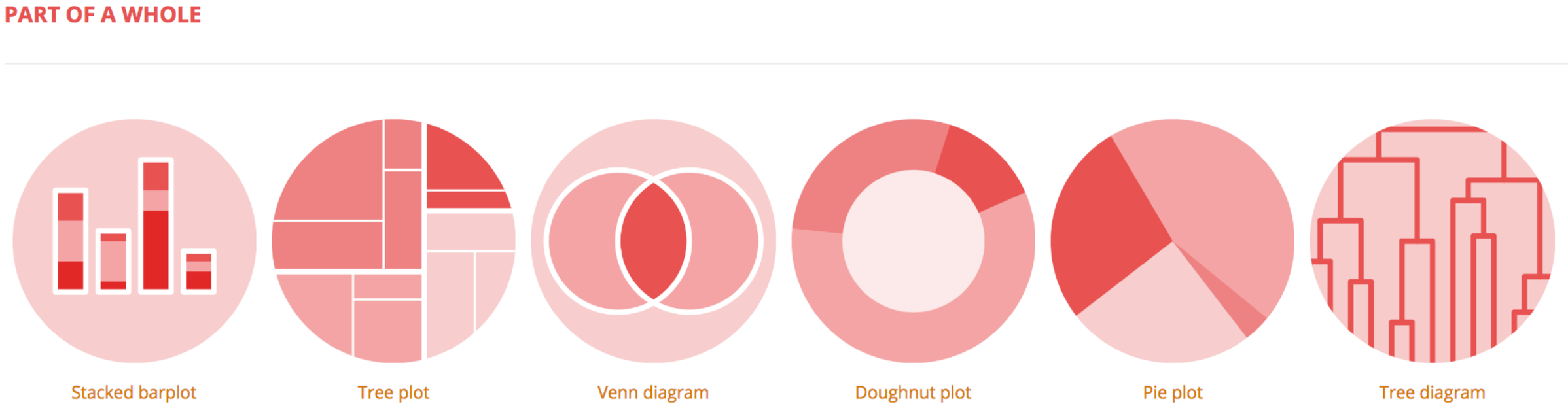
https://www.youtube.com/watch?v=E91bGT9BjYk

What chart for what scenario?

- **Evolution** – changes over time.

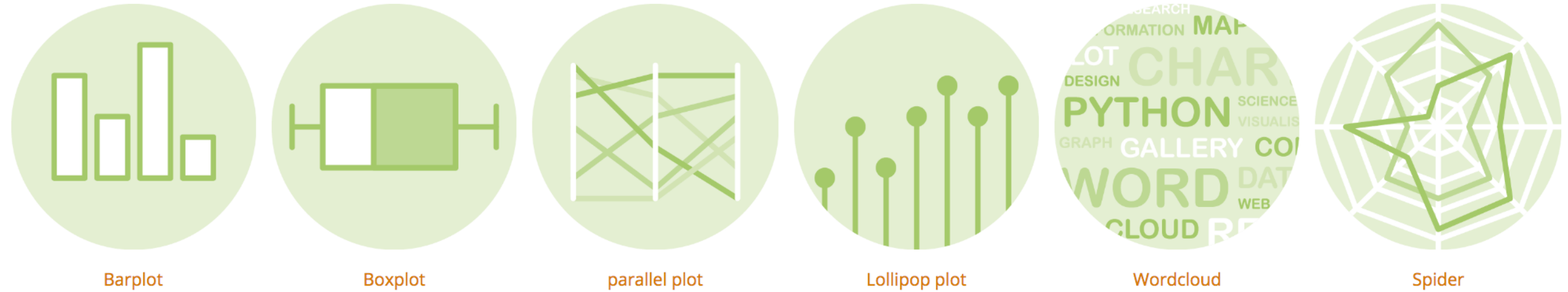
What chart for what scenario?

- **Parts of a whole** – Individual parts of a total.



What chart for what scenario?

- **Ranking** – Importance among categories.



- https://python-graph-gallery.com/

TOOLS

- So you have developers you can ask to do this for you…

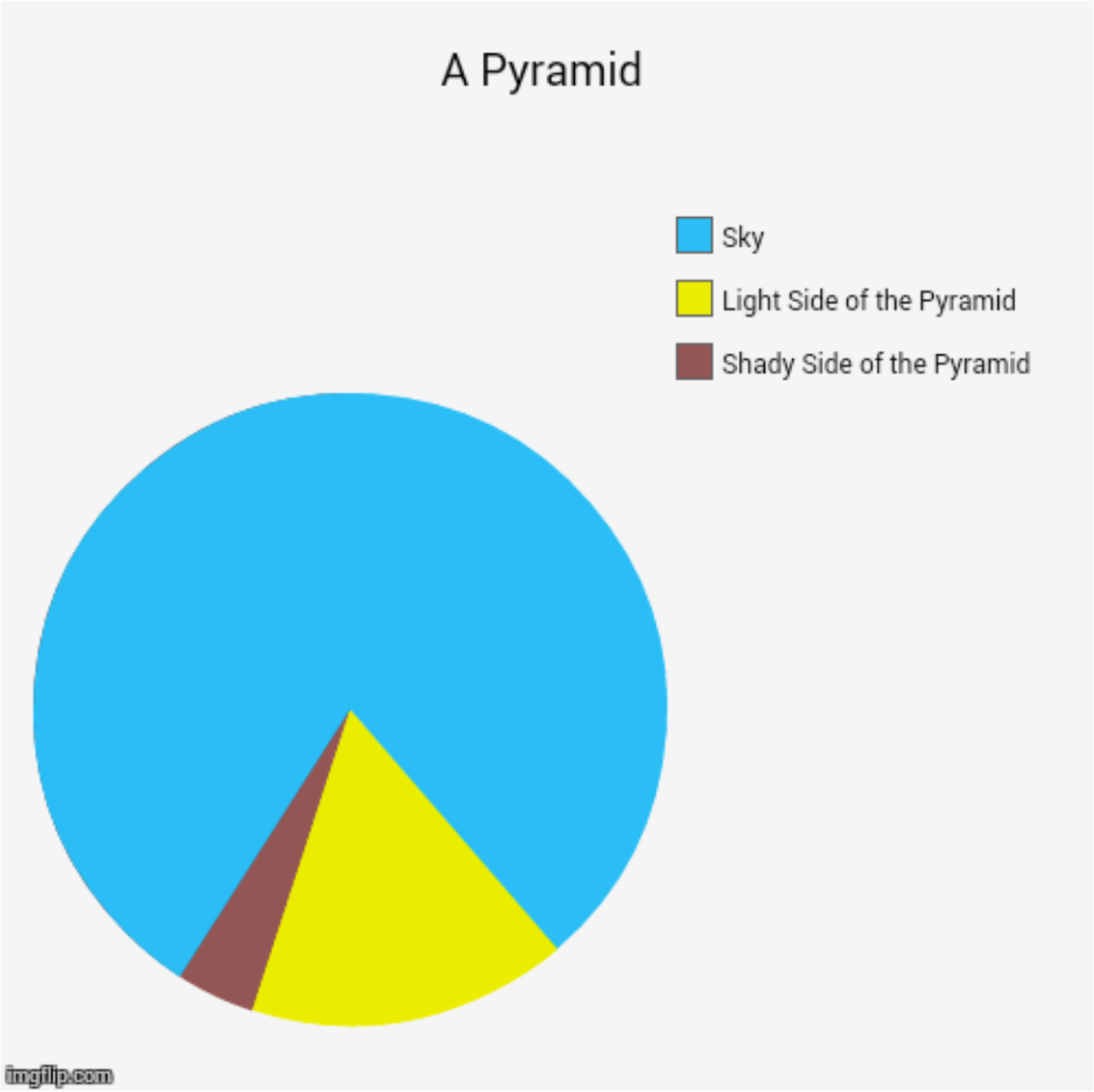
- D3.js: https://flowingdata.com/2015/12/15/a-day-in-the-life-of-americans/
- Tableau: https://public.tableau.com/en-us/s/blog/2015/07/analyzing-airbnb-data

- Excel + PPT (for real!)
- Canva
- Piktochart
- Python (Bokeh, Seaborn, matplotlib)
- R (ggplot)

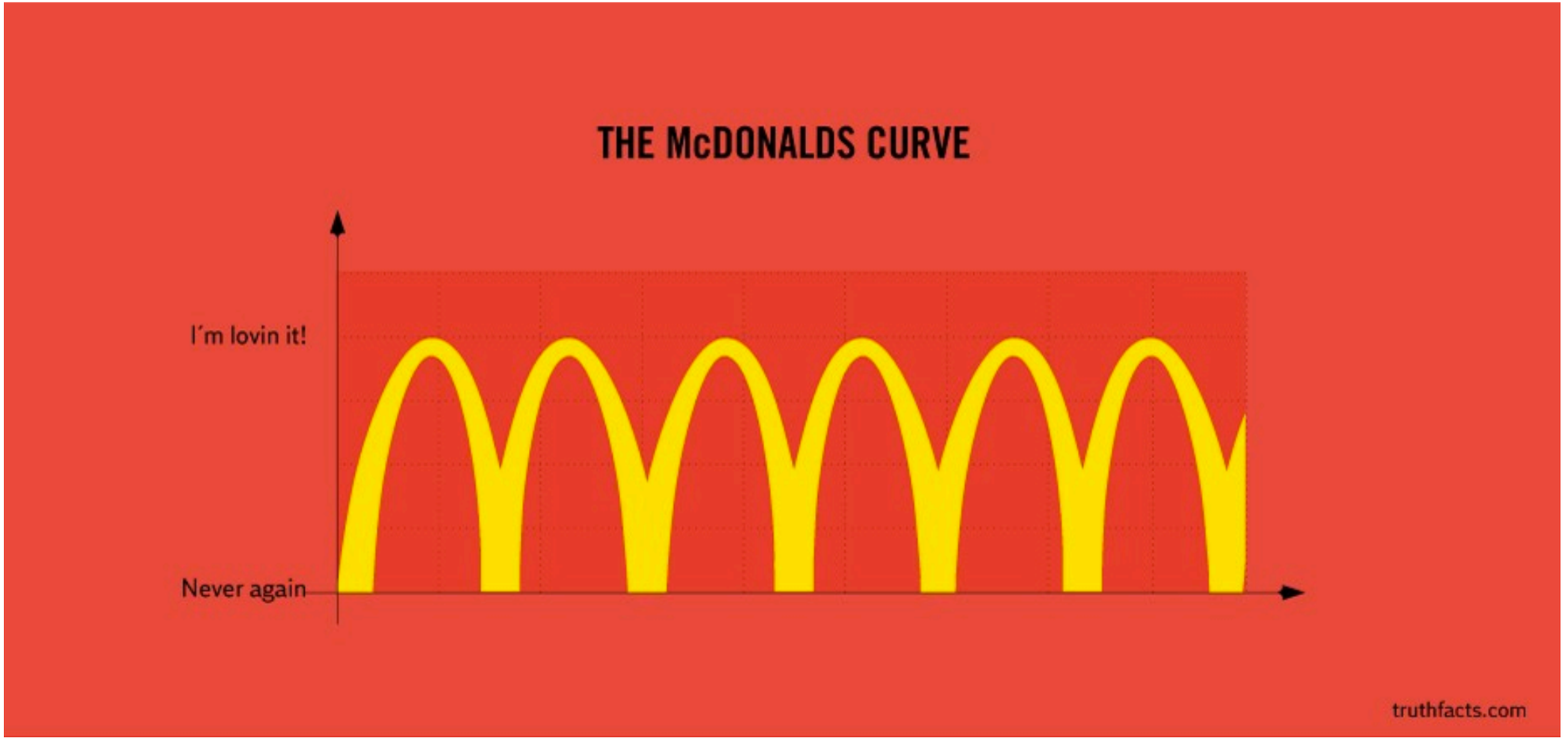
CHART JUNK

(JUST FOR LAUGHS)

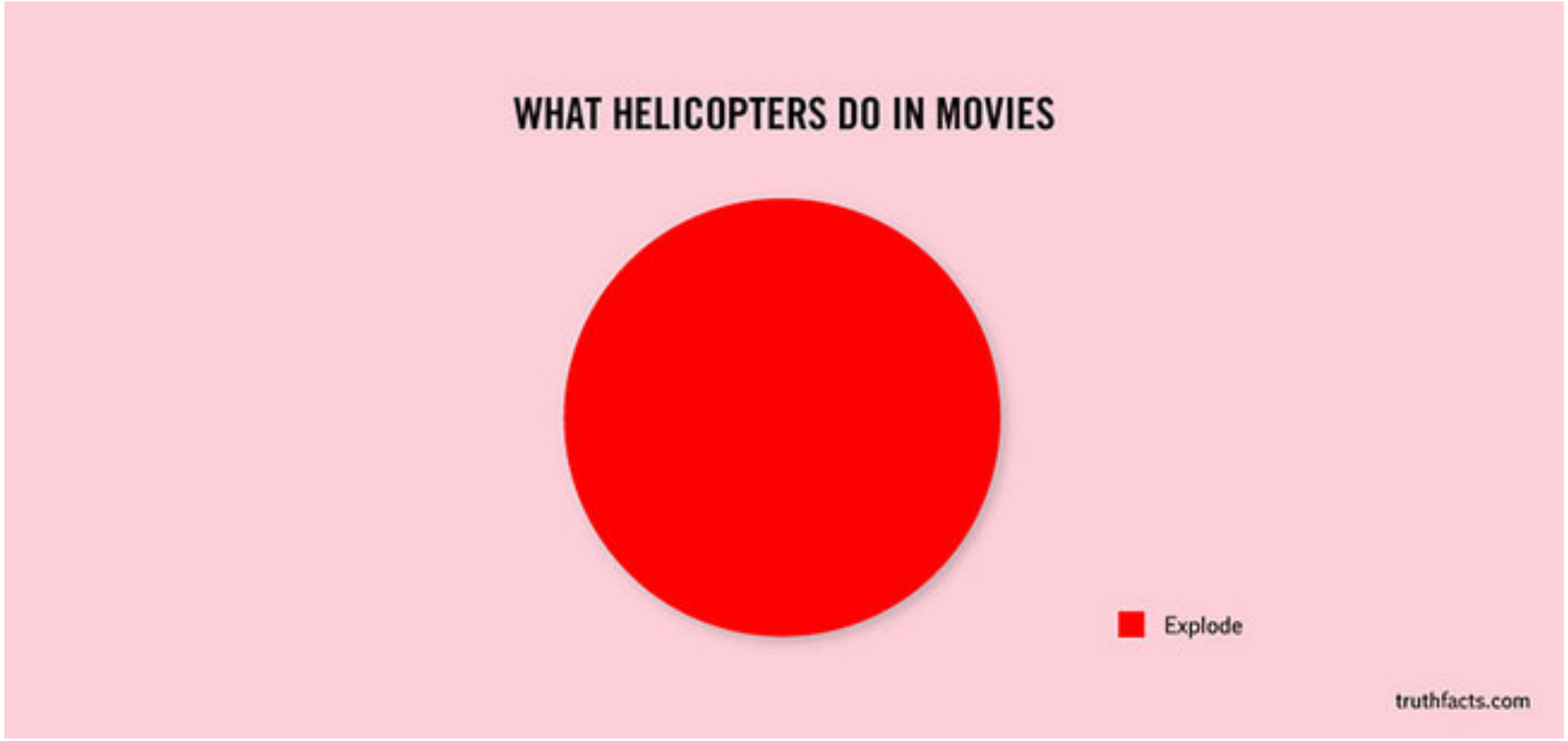
CHARTJUNK


A Pyramid
Sky
Light Side of the Pyramid
Shady Side of the Pyramid
imgflip.com

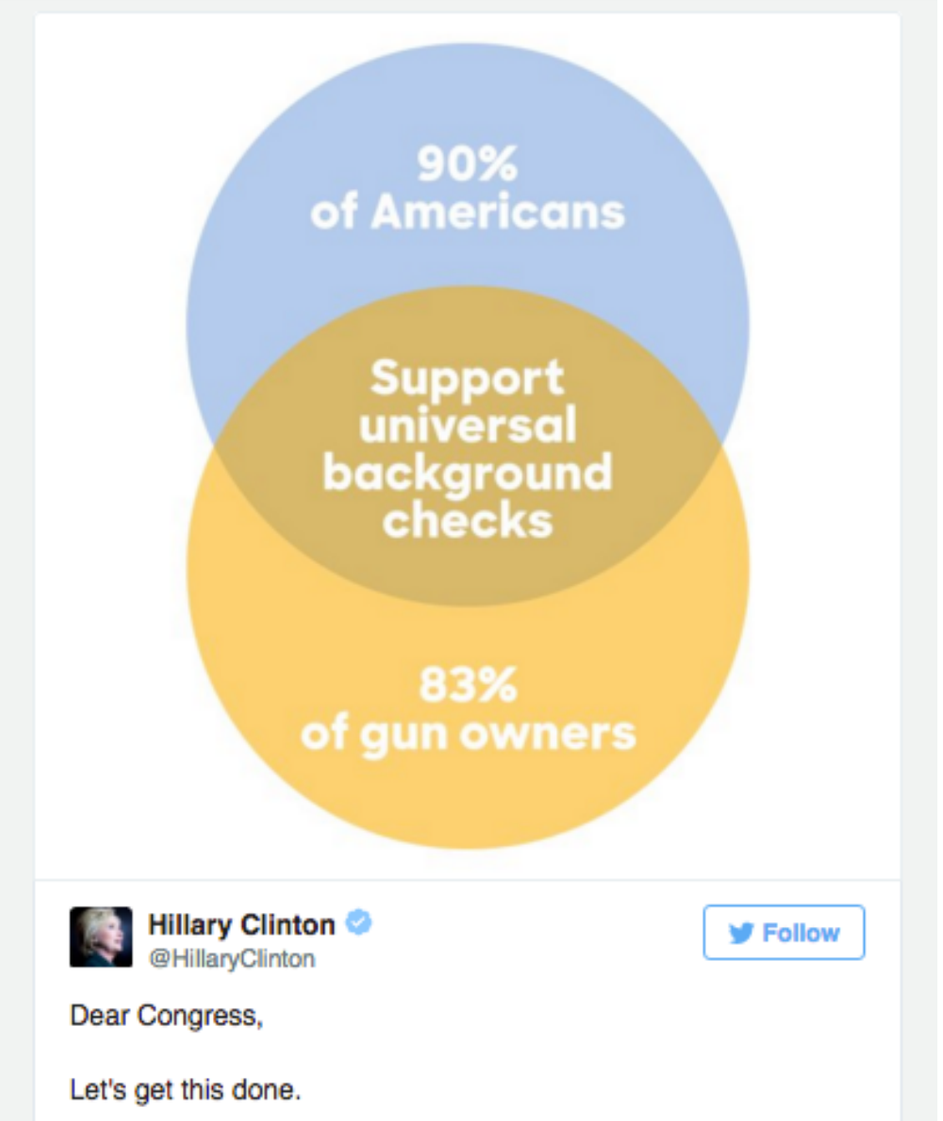
CHARTJUNK



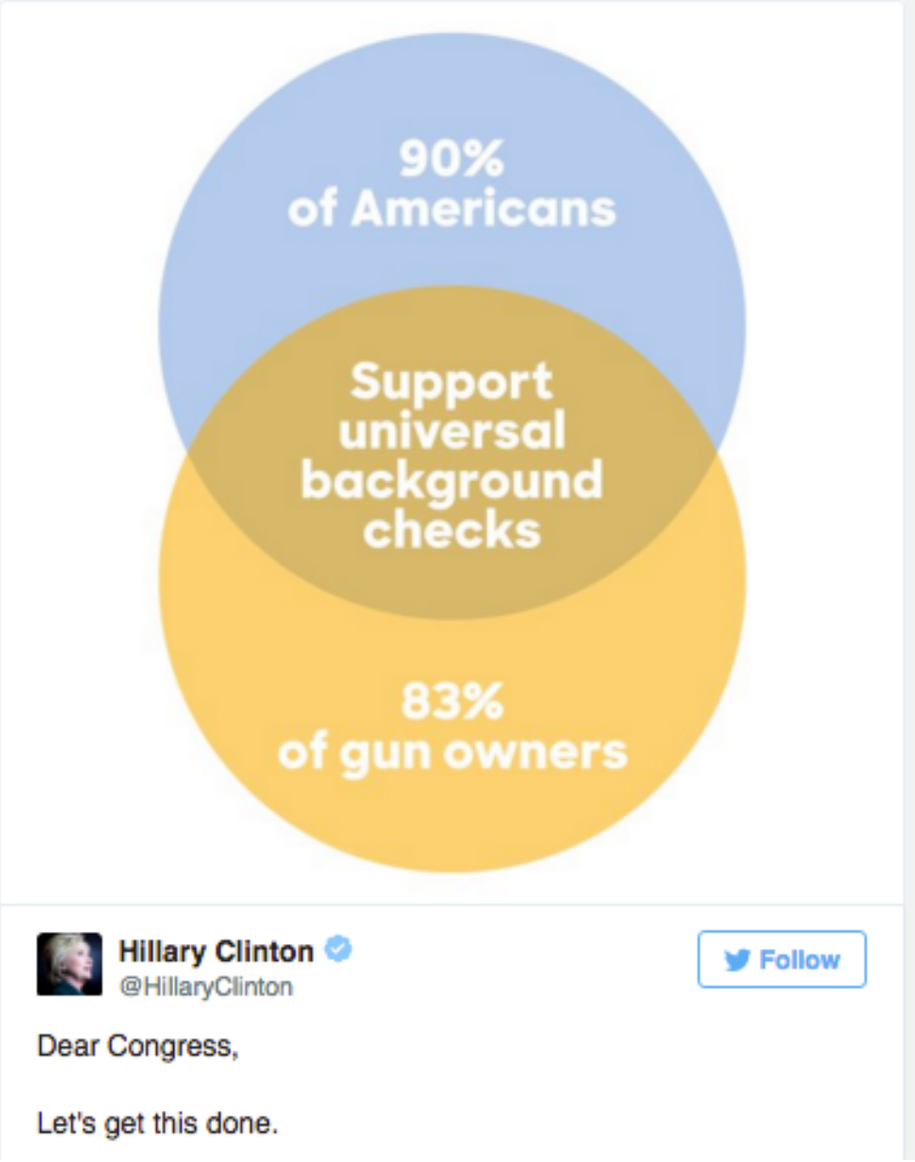
CHARTJUNK



GRAPHJUNK


90%
of Americans
Support
universal
background
checks
83%
of gun owners
Hillary Clinton
@HillaryClinton
Follow
Dear Congress,
Let's get this done.

GRAPHJUNK


90%
of Americans
Support
universal
background
checks
83%
of gun owners
Hillary Clinton
@HillaryClinton
Follow
Dear Congress,
Let's get this done.


People who know
how to make
Venn Diagrams
Hillary's graphic
design staff
In reply to Hillary Clinton
Michael Deppisch @deppisch · May 20
.@hillaryclinton
3.7K
5.4K

IT'S A BIPARTISAN PROBLEM

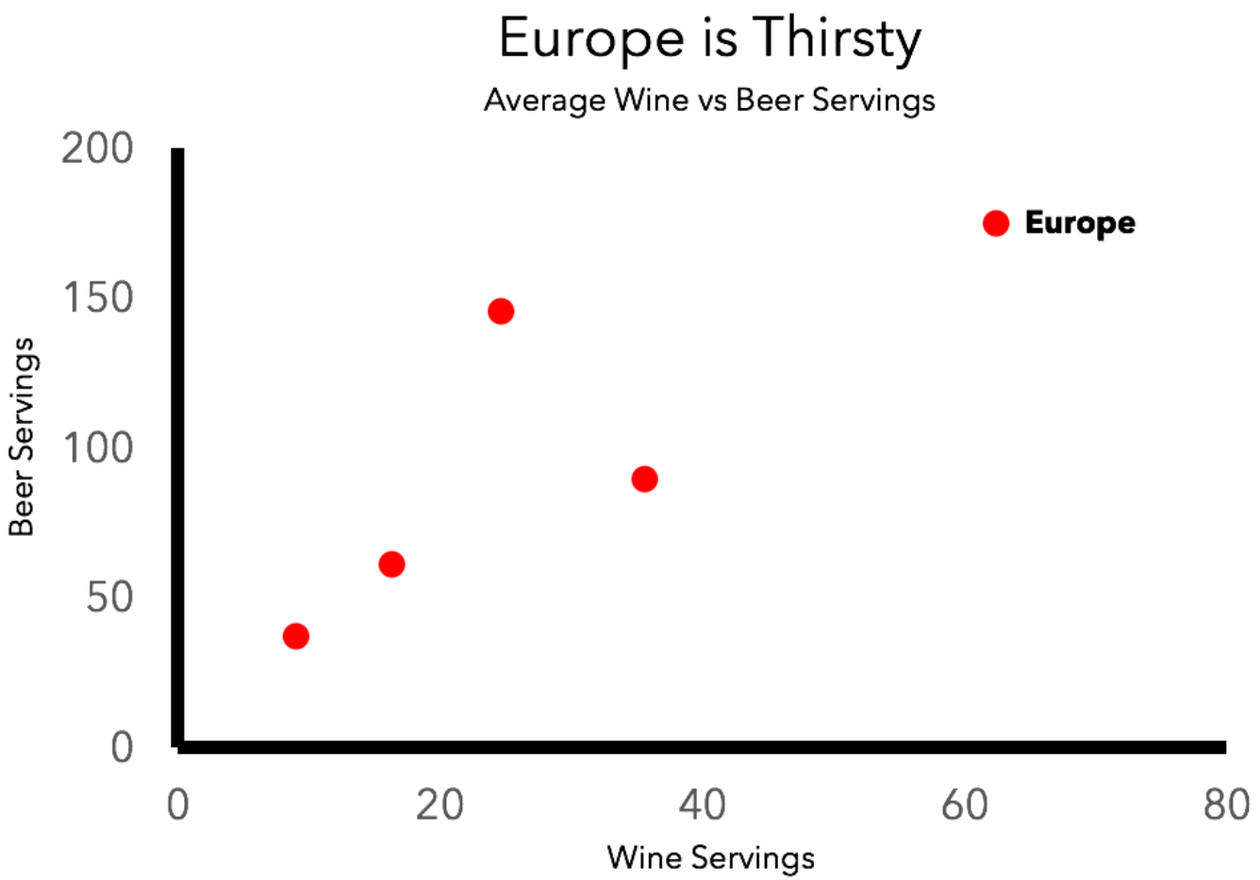
EXERCISE TIME!

Use the data included in the Google Folder

- Create an intriguing visualization summarizing the world's drinking habits.
- First: identify a narrative (to challenge or prove)
- Second: Identify the correct type of graph to demonstrate that narrative.
- Third: Format the data to match our desired outcomes.
- Fourth: Use PowerPoint to improve our visualization's appearance.

Let’s make one together, first…

My example!


Europe is Thirsty
Average Wine vs Beer Servings
200
150
100
50
0
Beer Servings
0
20
40
60
80
Wine Servings
Europe

GENERAL ASSEMBLY

THANK YOU

QUESTIONS, COMMENTS

@josephofiowa
joseph@represently.com