

Visualization, EVA, & Uncertainty

A Crash Course

Arvind Satyanarayan

LES VARIABLES DE SÉPARATION DES IMAGES

GRAIN



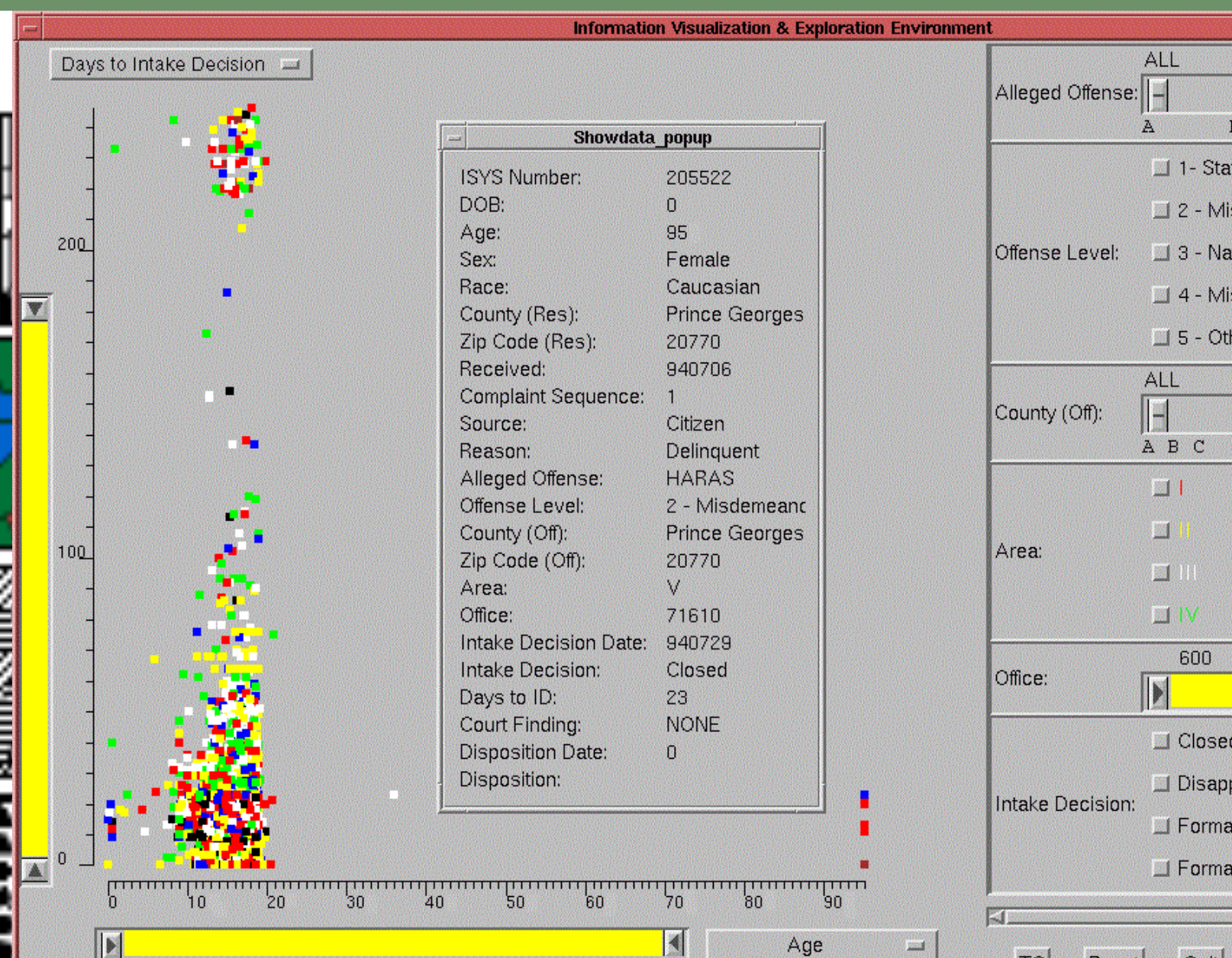
COULEUR



ORIENTATION



FORME

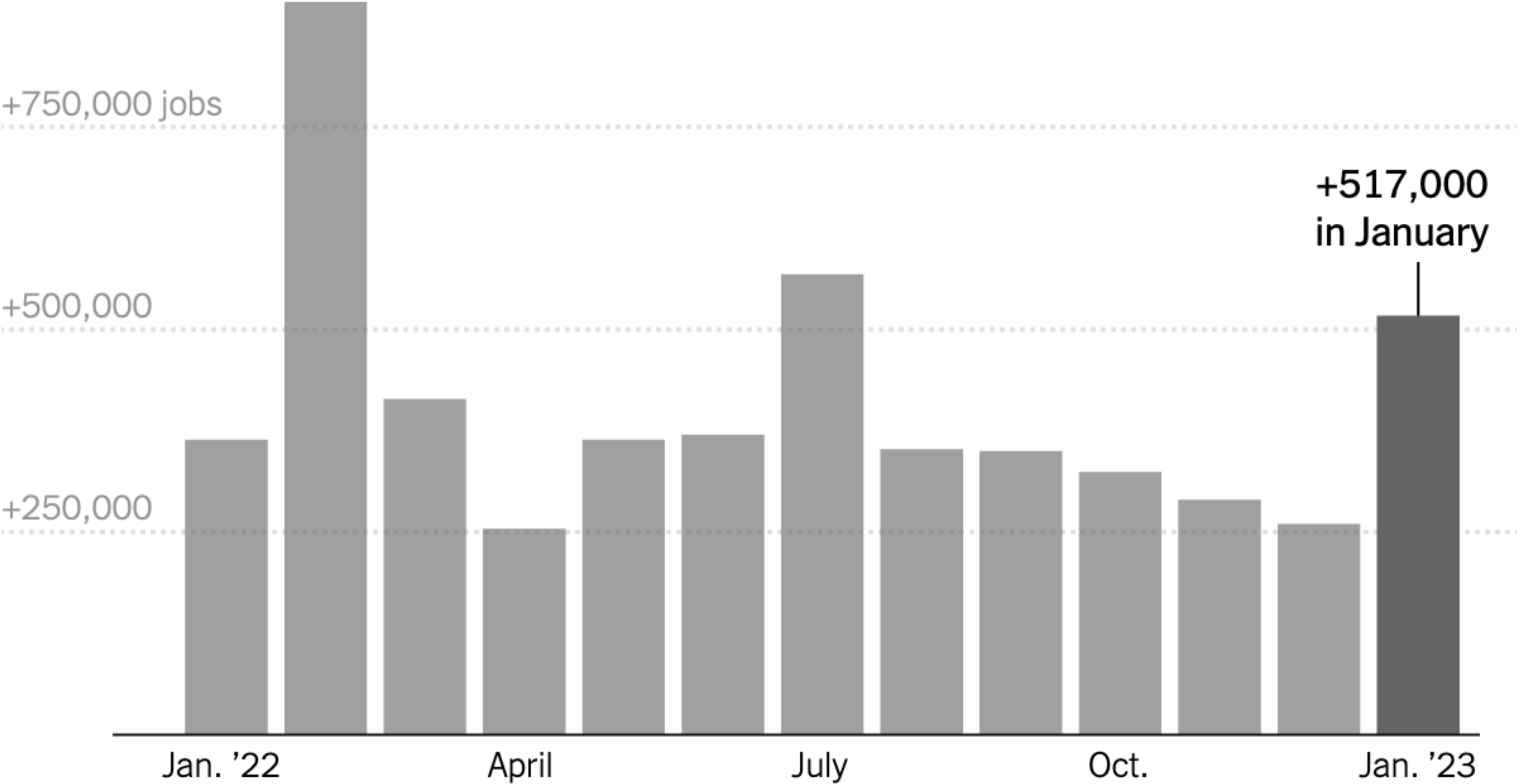


PART ONE

Data & Image Models

Name That Chart


Monthly change in jobs



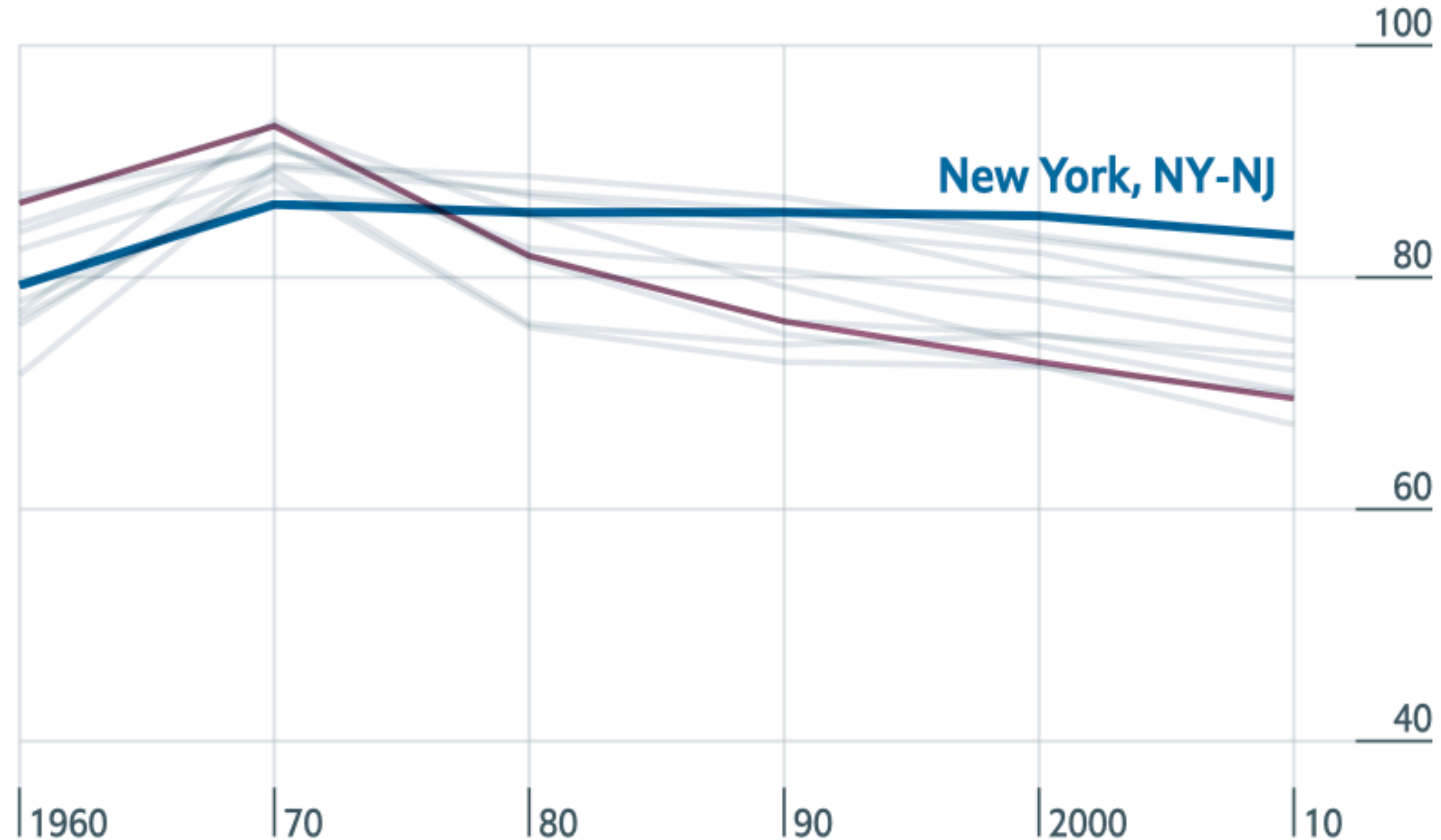
Data is seasonally adjusted. • Source: Bureau of Labor Statistics • By Ella Koeze

Black-white segregation in 60 biggest metro areas

— United States average

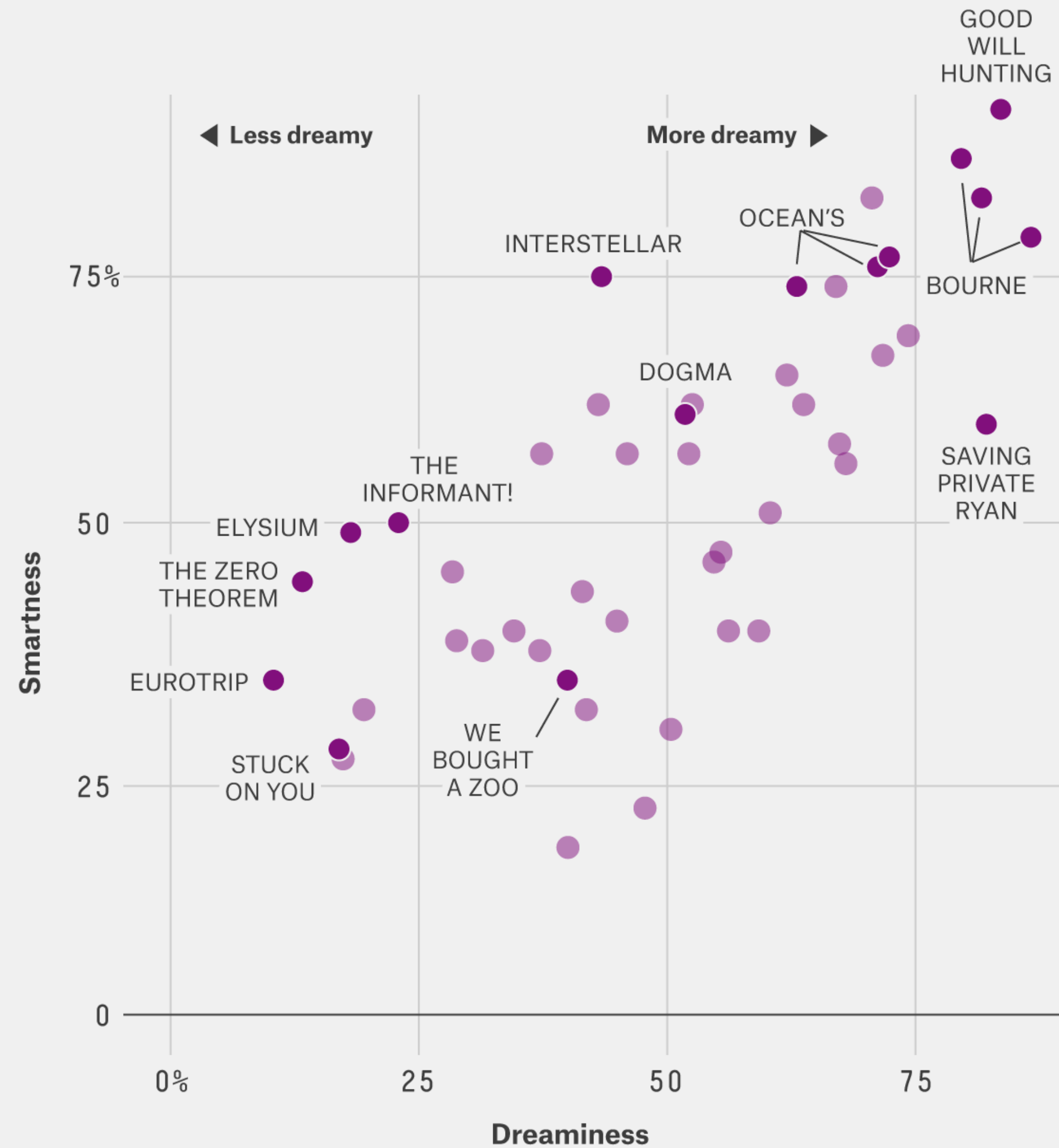
Northeast 

Complete segregation



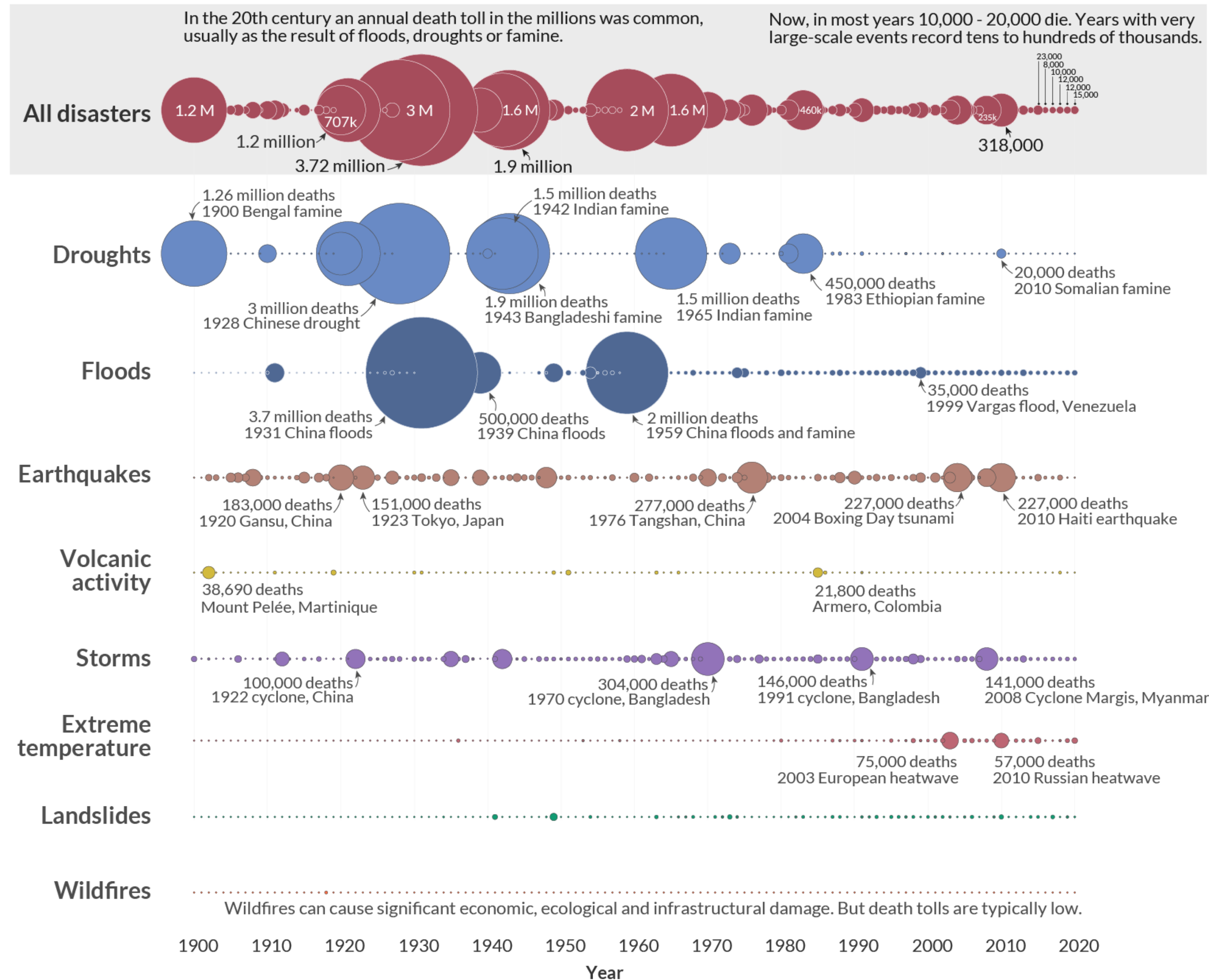
Matt Damon Is Dreamy Whenever He Is Smart

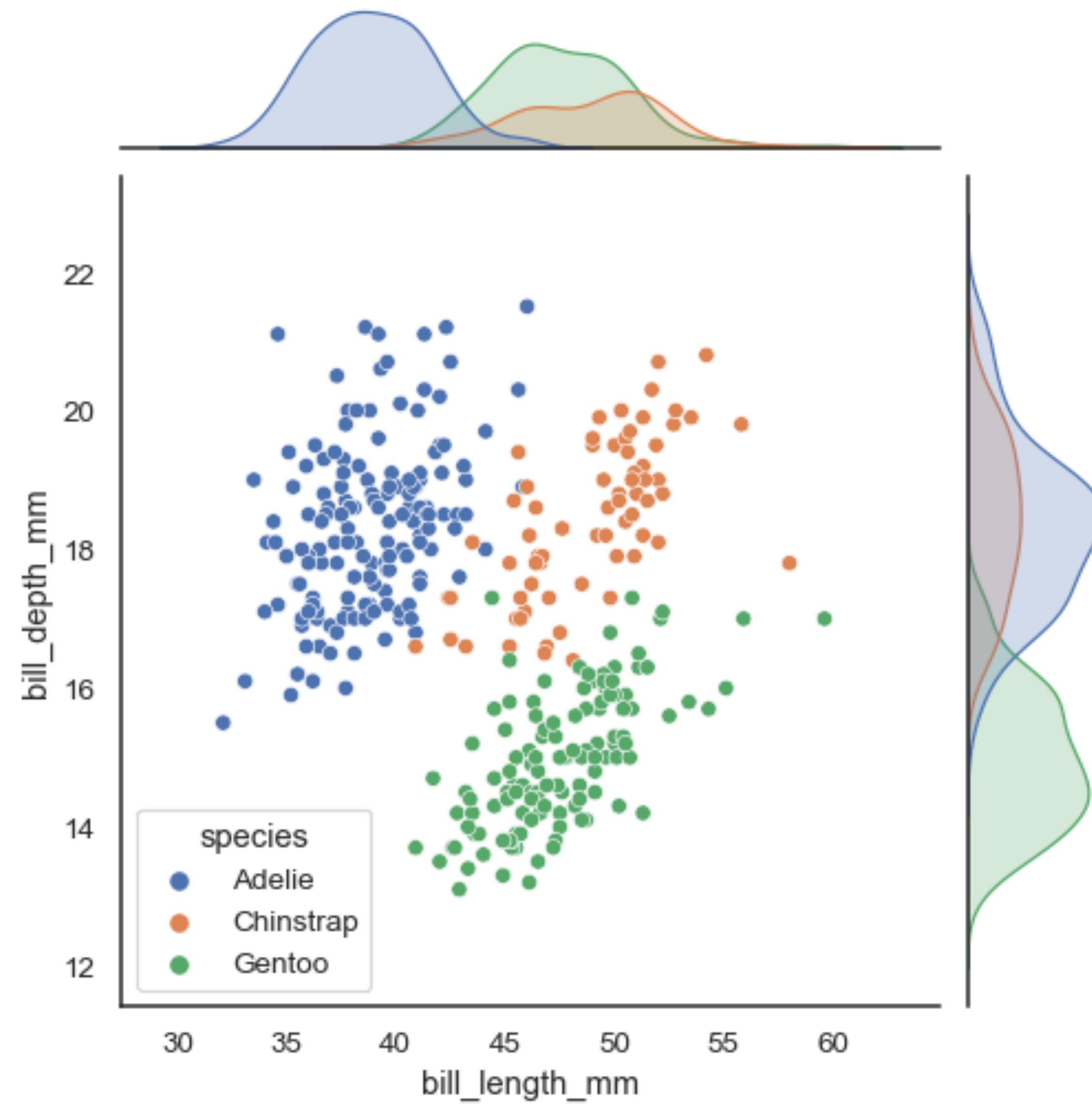
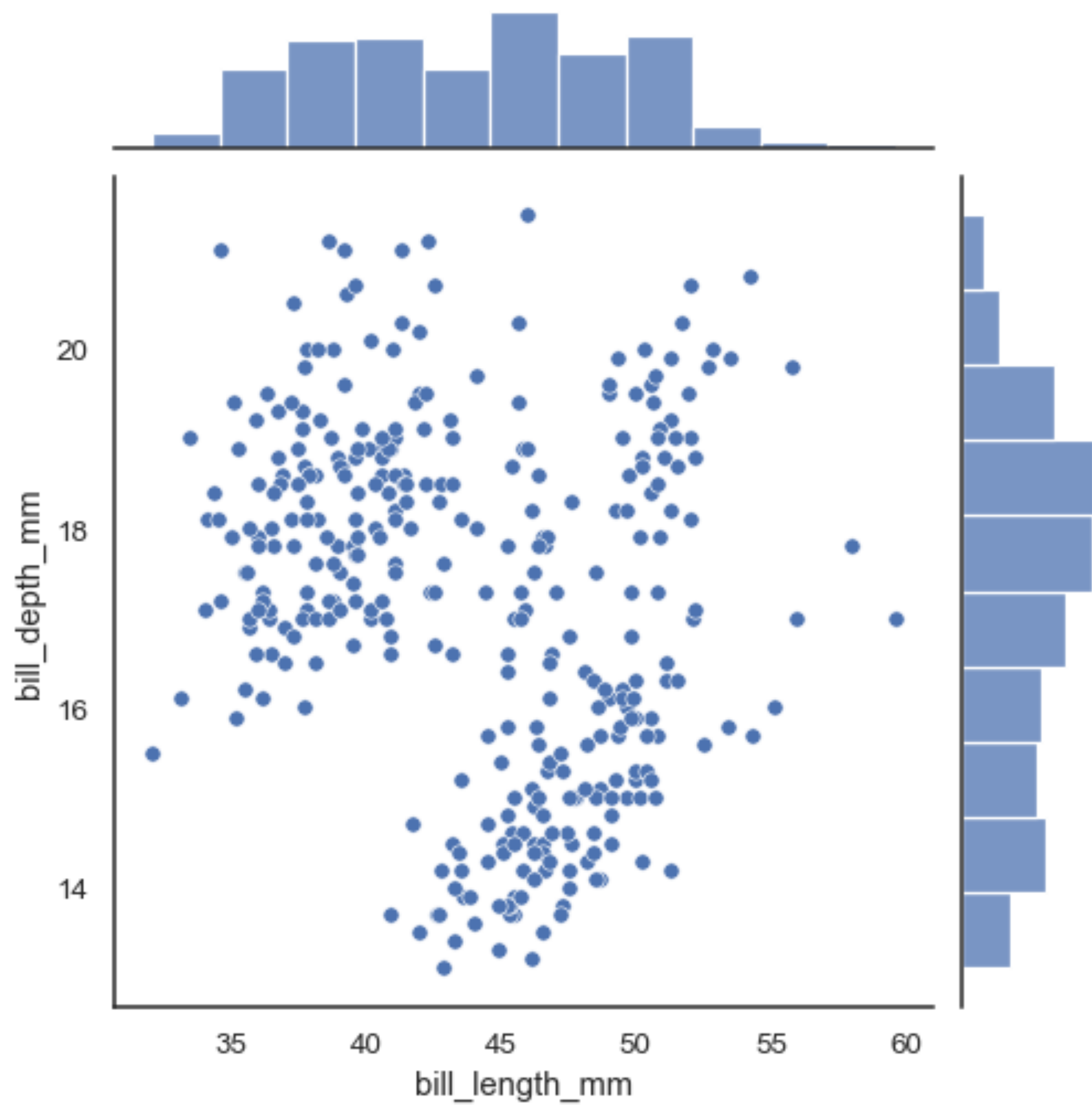
“Smartness” vs. “dreaminess” based on probabilities that a character played by Matt Damon will beat an average Matt Damon in the category, from surveys of 3,435 respondents about the smartness of characters and 17,582 about the dreaminess



Global deaths from disasters over more than a century

The size of the bubble represents the estimated annual death toll. The largest years are labeled with this total figure, alongside large-scale events that contributed to the majority – although usually not all – of these deaths.





Driving Shifts Into Reverse

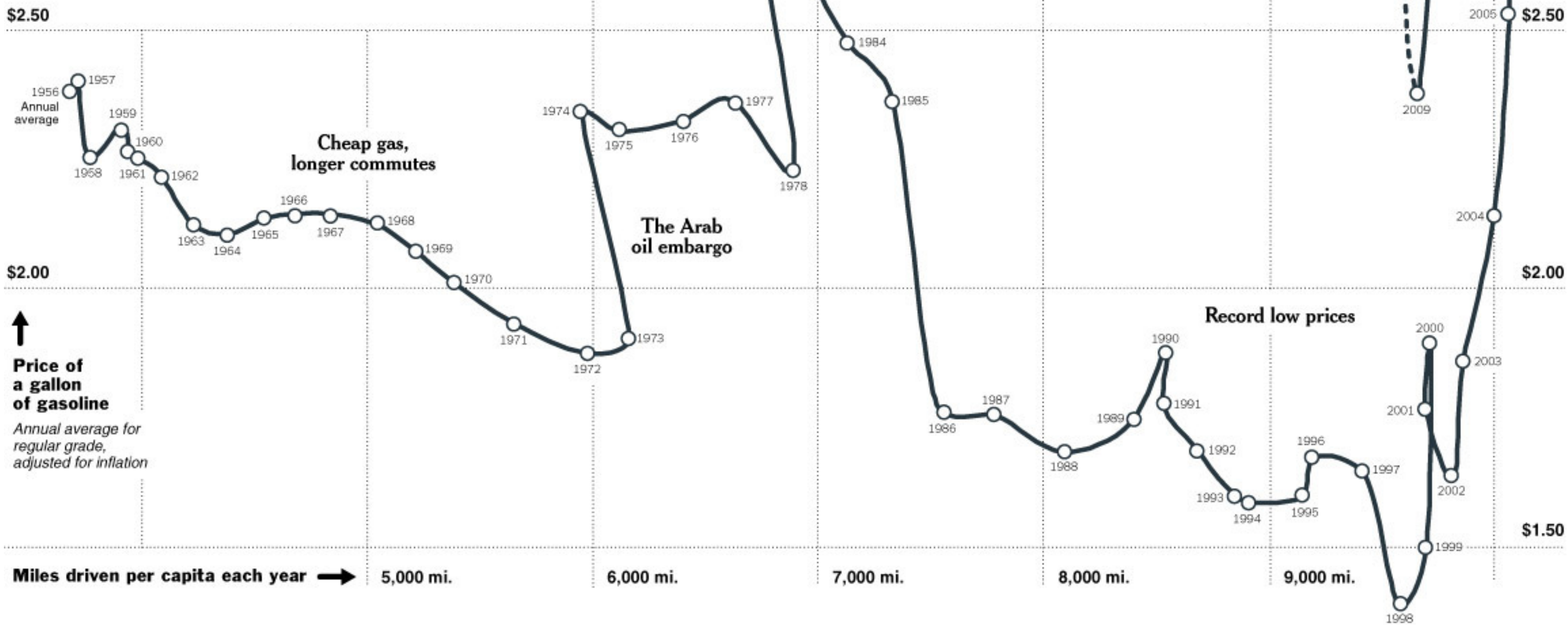
ECONOMISTS have long studied the relationship between driving habits and gasoline prices. Low gas prices can bring periods of profligate driving, and a quick jump in prices can cause many vehicles to languish in garages.

Until recently, Americans have driven more each year than the previous one, with a few brief exceptions. In 1956, Americans of driving age drove about 4,000 miles a year, on average. Fifty years later, that figure had climbed above 10,000.

But the latest recession has caused some big changes. High unemployment meant that fewer people were driving to work, and a slump in consumer spending

meant that less freight needed to be moved around the country. As gas prices soared in 2005, the number of miles driven — including commercial and personal — began to fall, and continued to drop after 2008 even as gasoline became cheaper.

“People were surprised by the very rapid rise in gas prices, and they changed their driving behavior,” said Kenneth A. Small, a transportation economist at the University of California, Irvine. “But my suspicion is that it is temporary. As soon as unemployment gets back to pre-recession levels, we will see Americans doing a lot more driving again.”



Recommended Charts

Maps PivotChart Sparklines Slicer Timeline

Combo

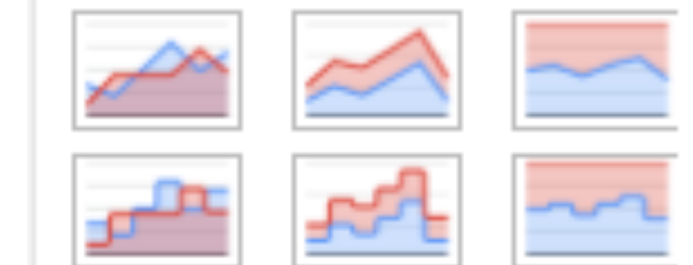
Chart type

Pie chart

Line



Area



Column



Bar



Pie



Scatter



Map



Other



- seaborn.relplot
- seaborn.scatterplot
- seaborn.lineplot
- seaborn.displot
- seaborn.histplot
- seaborn.kdeplot
- seaborn.ecdfplot
- seaborn.rugplot
- seaborn.distplot
- seaborn.catplot
- seaborn.stripplot
- seaborn.swarmplot
- seaborn.boxplot
- seaborn.violinplot
- seaborn.boxenplot
- seaborn.pointplot
- seaborn.barplot
- seaborn.countplot
- seaborn.lmplot
- seaborn.regplot
- seaborn.residplot
- seaborn.heatmap
- seaborn.clustermap
- seaborn.FacetGrid
- seaborn.pairplot
- seaborn.PairGrid
- seaborn.jointplot
- seaborn.JointGrid

Data Visualization

Data

Mapping or Visual Encoding



Visual

Attribute Types

Attribute Types

Nominal

Labels or categories.

E.g., Fruits: apples, bananas, cantaloupes, ...

Attribute Types

Nominal

Labels or categories.

E.g., Fruits: apples, bananas, cantaloupes, ...

Ordinal

Ordered.

E.g., Quality of meat: Grade A, AA, AAA

Attribute Types

Nominal

Labels or categories.

E.g., Fruits: apples, bananas, cantaloupes, ...

Ordinal

Ordered.

E.g., Quality of meat: Grade A, AA, AAA

Quantitative (Interval)

Interval (zero can be arbitrarily located).

E.g., Dates: Jan 19, 2018; Location: (Lat 42.36, -71.09)

Only differences can be calculated (e.g., distances or spans).

Quantitative (Ratio)

Attribute Types

Nominal

Labels or categories.

E.g., Fruits: apples, bananas, cantaloupes, ...

Ordinal

Ordered.

E.g., Quality of meat: Grade A, AA, AAA

Quantitative (Interval)

Interval (zero can be arbitrarily located).

E.g., Dates: Jan 19, 2018; Location: (Lat 42.36, -71.09)

Only differences can be calculated (e.g., distances or spans).

Quantitative (Ratio)

Ratio (fixed zero / meaningful baseline).

E.g., Physical measurement: length, mass, temperature

Counts and amounts. Can measure ratios or proportions.

Attribute Types

Nominal

=, ≠

Labels or categories.

E.g., Fruits: apples, bananas, cantaloupes, ...

Ordinal

=, ≠, <, >

Ordered.

E.g., Quality of meat: Grade A, AA, AAA

Quantitative (Interval)

=, ≠, <, >, -

Interval (zero can be arbitrarily located).

E.g., Dates: Jan 19, 2018; Location: (Lat 42.36, -71.09)

Only differences can be calculated (e.g., distances or spans).

Quantitative (Ratio)

=, ≠, <, >, -, %

Ratio (fixed zero / meaningful baseline).

E.g., Physical measurement: length, mass, temperature

Counts and amounts. Can measure ratios or proportions.

Visual Variables

Also called visual *channels*.

Used to encode data values as characteristics of marks.

** From 1967, so Bertin only accounted for visualizations that were printable, white paper.*

LES VARIABLES DE L'IMAGE

	POINTS			LIGNES			ZONES	
XY 2 DIMENSIONS DU PLAN	x	x	x	/	~	/	14 1 18 21 2 14 15 1	2 18 1 21 15 1 2 9
Z TAILLE	█	█	█	/	~	/	█	█
VALEUR	█	█	█	/	~	/	█	█

LES VARIABLES DE SÉPARATION DES IMAGES

GRAIN	█	█	█	/	~	/	█	█
COULEUR	█	█	█	/	~	/	█	█
ORIENTATION	█	█	█	/	~	/	█	█
FORME	█	█	█	/	~	/	█	█

Marks

Basic graphical elements that represent data items.



Channels: Expressiveness Types and Effectiveness Ranks

➔ Magnitude Channels

Position on common scale



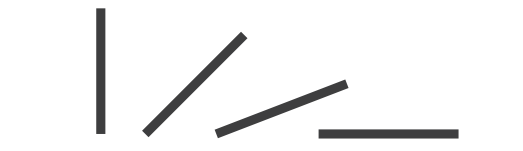
Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



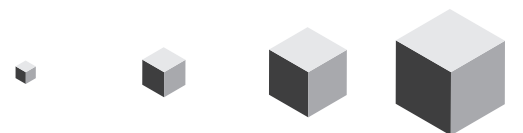
Color saturation



Curvature



Volume (3D size)



➔ Identity Channels

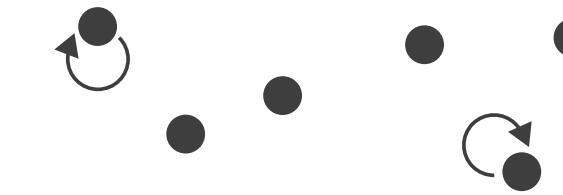
Spatial region



Color hue



Motion



Shape



Tamara Munzner, *Visualization Analysis and Design* (2014).

Channels: Expressiveness Types and Effectiveness Ranks

➔ **Magnitude Channels: O or Q** attributes

Position on common scale



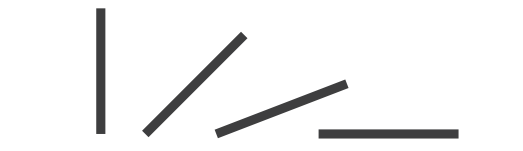
Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



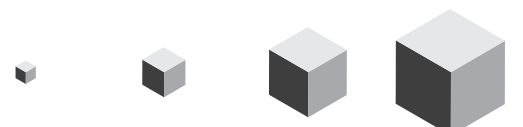
Color saturation



Curvature



Volume (3D size)



➔ **Identity Channels: N** attributes

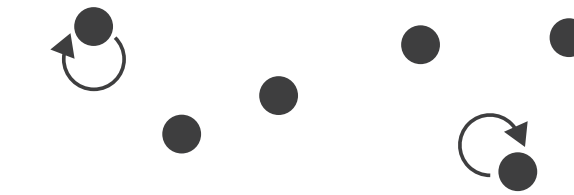
Spatial region



Color hue



Motion



Shape



Tamara Munzner, *Visualization Analysis and Design* (2014).

Channels: Expressiveness Types and Effectiveness Ranks

➔ Magnitude Channels: O or Q attributes

Position on common scale



Position on unaligned scale



Length (1D size)



Tilt/angle



Area (2D size)



Depth (3D position)



Color luminance



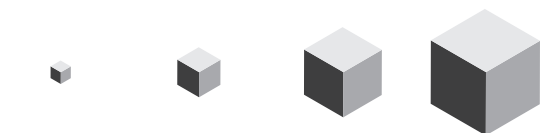
Color saturation



Curvature



Volume (3D size)



Same

Same

Same

Most Effectiveness Least

➔ Identity Channels: N attributes

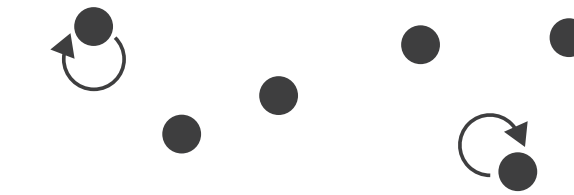
Spatial region



Color hue



Motion



Shape

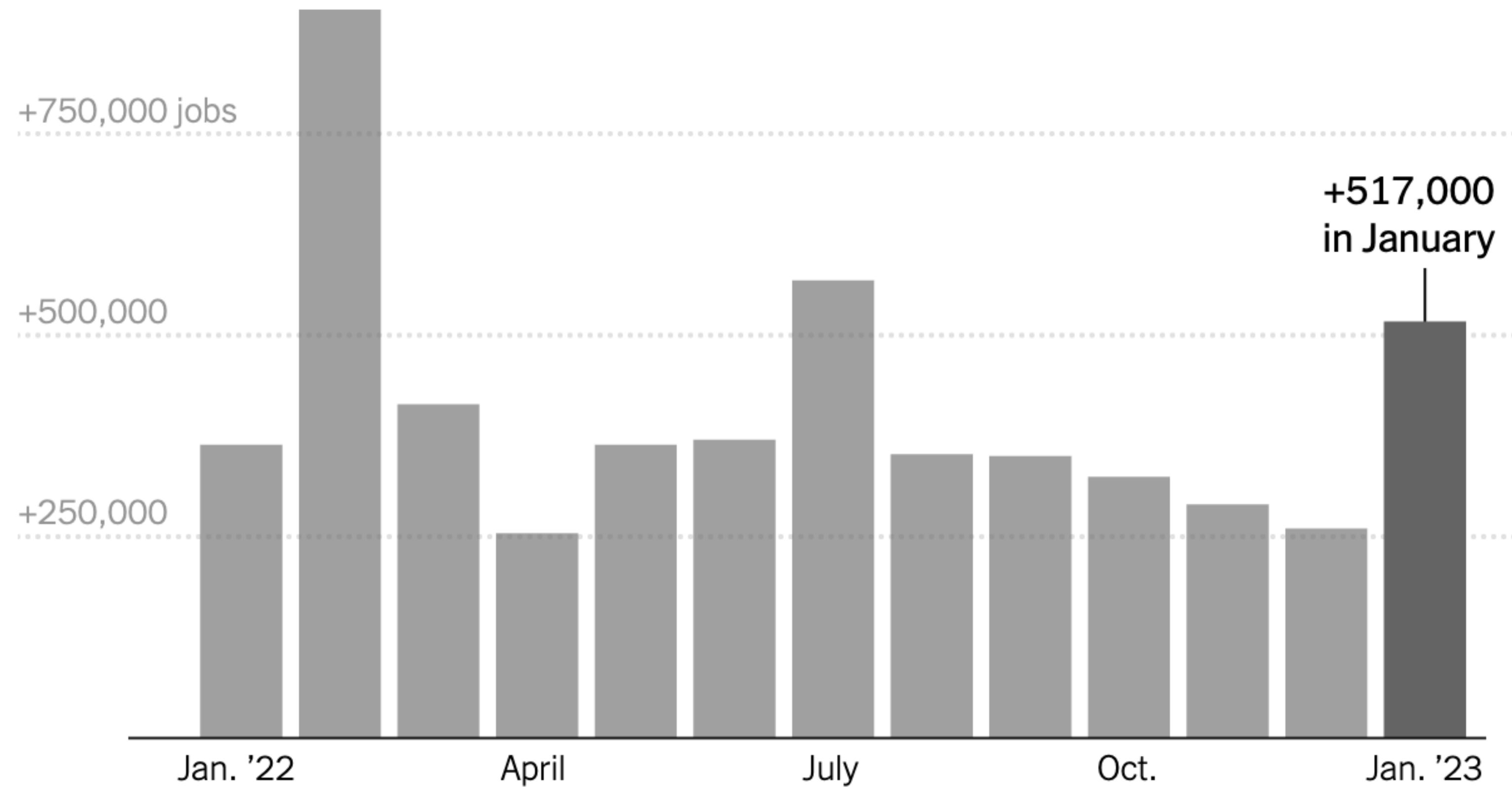


Tamara Munzner, *Visualization Analysis and Design* (2014).

Name That Chart

*Visual
Encoding*

Monthly change in jobs



Data is seasonally adjusted. • Source: Bureau of Labor Statistics • By Ella Koeze

Mark: bar

X-Axis: month (O)

Y-Axis: jobs (Q-Ratio)

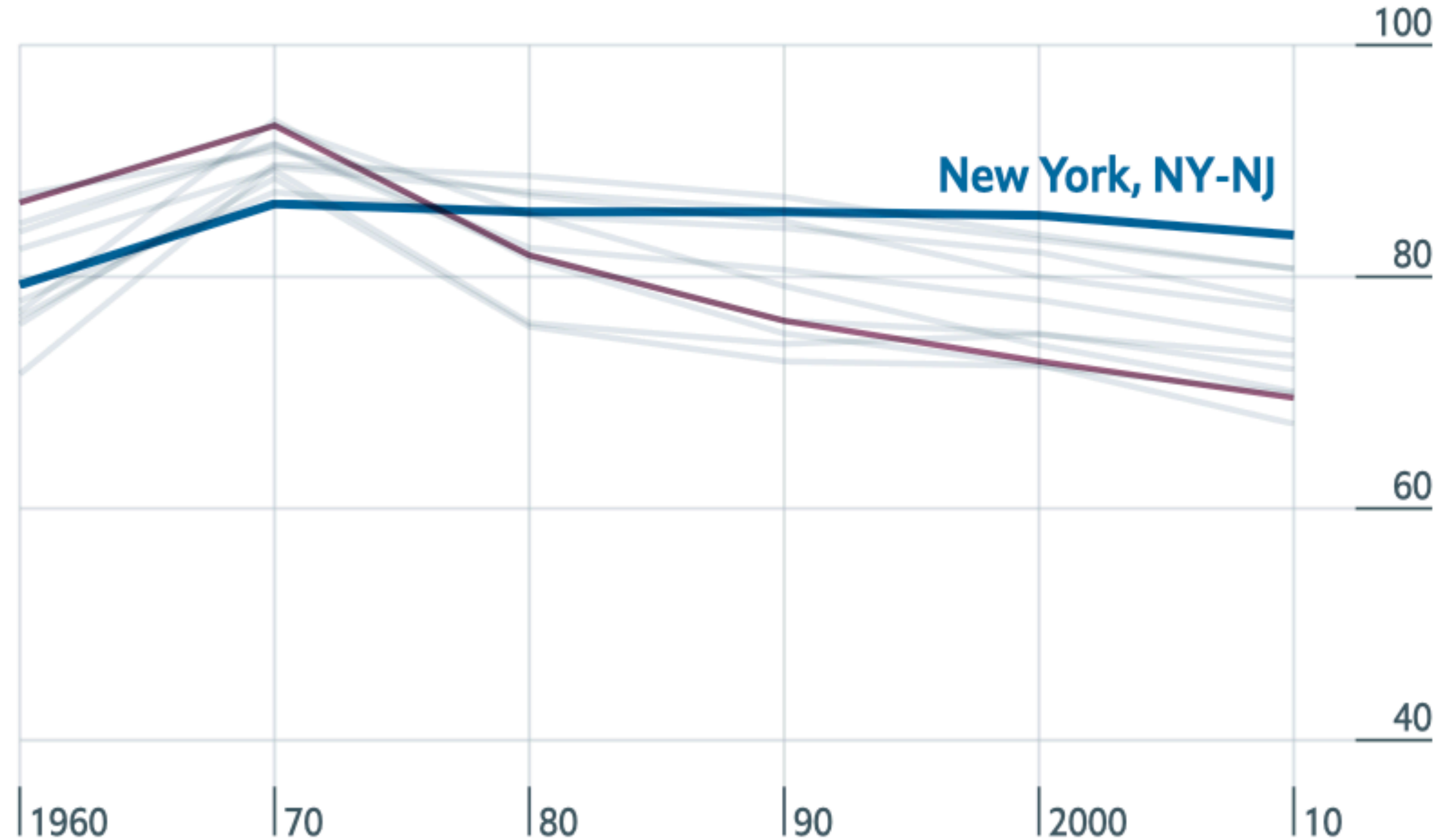
Black-white segregation in 60 biggest metro areas

— United States average

Northeast



Complete segregation



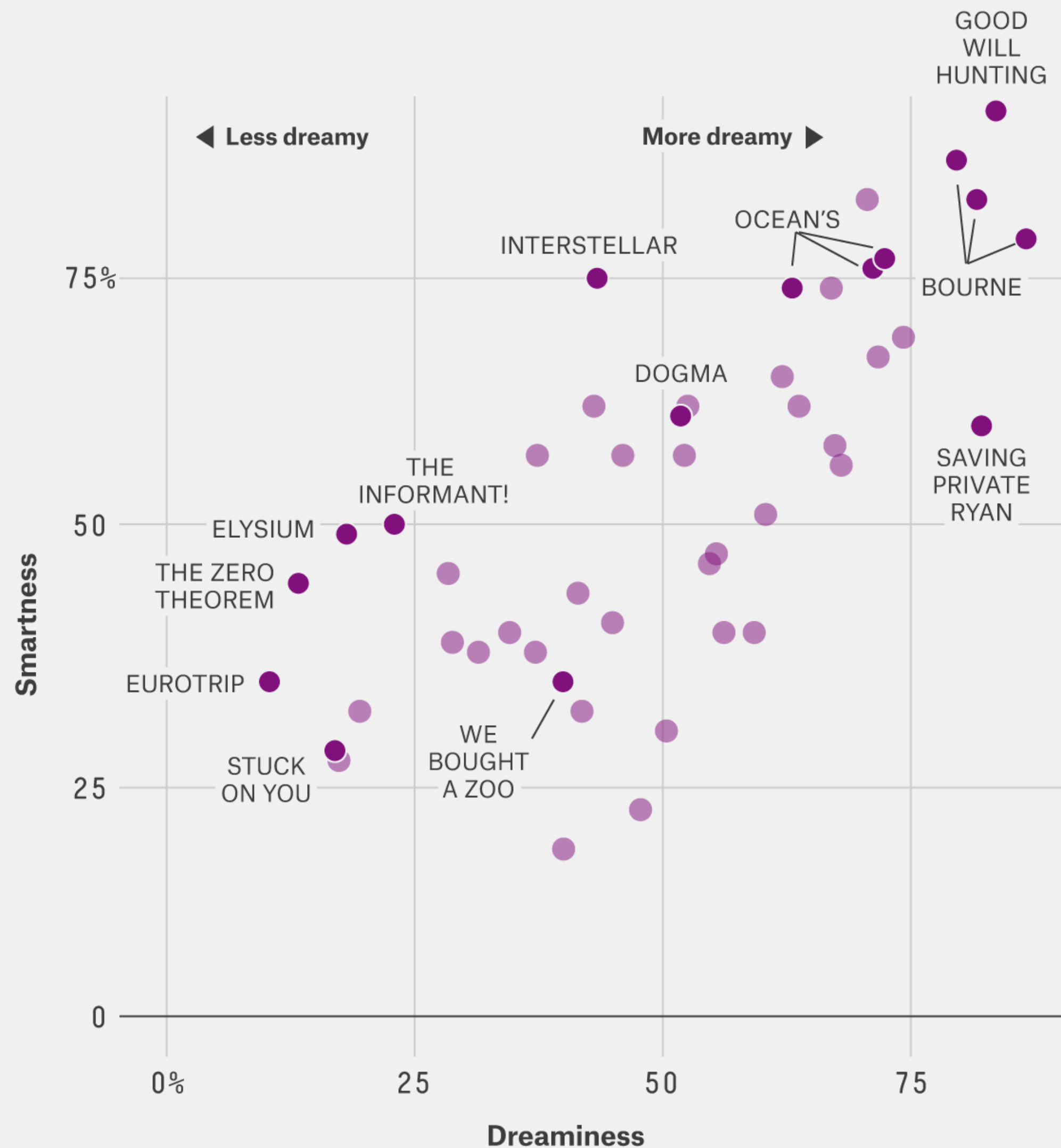
Mark: line

X-Axis: year (Q-Interval)

Y-Axis: segregation index (Q-Ratio)

Matt Damon Is Dreamy Whenever He Is Smart

“Smartness” vs. “dreaminess” based on probabilities that a character played by Matt Damon will beat an average Matt Damon in the category, from surveys of 3,435 respondents about the smartness of characters and 17,582 about the dreaminess



Mark: circle

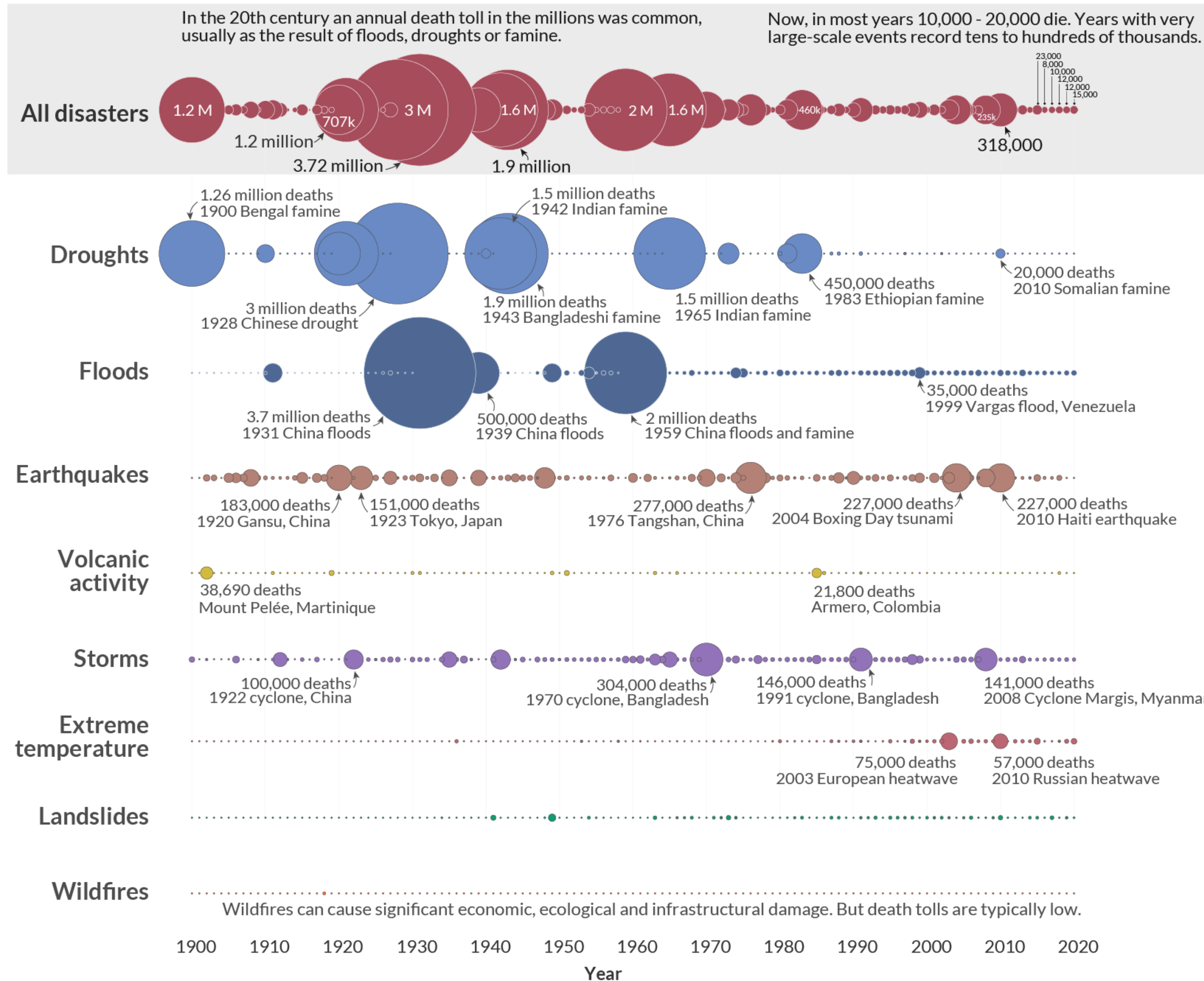
X-Axis: dreaminess (Q-Ratio)

Y-Axis: smartness (Q-Ratio)

Color: ?? (annotation purposes only)

Global deaths from disasters over more than a century

The size of the bubble represents the estimated annual death toll. The largest years are labeled with this total figure, alongside large-scale events that contributed to the majority – although usually not all – of these deaths.



Mark: circle

X-Axis: year (Ordinal)

Y-Axis: category (Nominal)

Color: category (Nominal)

Size: deaths (Q-Ratio)

METRICS
HANNAH FAIRFIELD

Driving Shifts Into Reverse

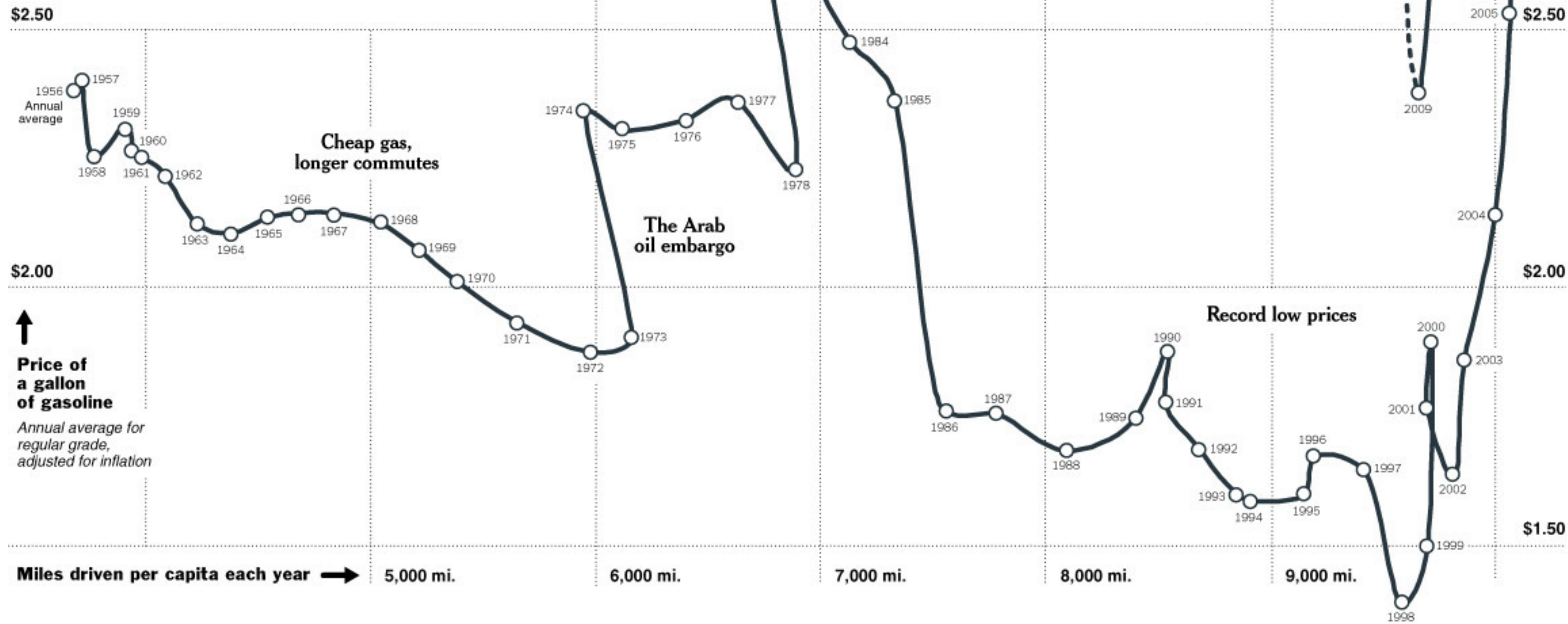
ECONOMISTS have long studied the relationship between driving habits and gasoline prices. Low gas prices can bring periods of profligate driving, and a quick jump in prices can cause many vehicles to languish in garages.

Until recently, Americans have driven more each year than the previous one, with a few brief exceptions. In 1956, Americans of driving age drove about 4,000 miles a year, on average. Fifty years later, that figure had climbed above 10,000.

But the latest recession has caused some big changes. High unemployment meant that fewer people were driving to work, and a slump in consumer spending

meant that less freight needed to be moved around the country. As gas prices soared in 2005, the number of miles driven — including commercial and personal — began to fall, and continued to drop after 2008 even as gasoline became cheaper.

“People were surprised by the very rapid rise in gas prices, and they changed their driving behavior,” said Kenneth A. Small, a transportation economist at the University of California, Irvine. “But my suspicion is that it is temporary. As soon as unemployment gets back to pre-recession levels, we will see Americans doing a lot more driving again.”



Mark: circle

X-Axis: miles (Q-Ratio)

Y-Axis: price (Q-Ratio)

+

Mark: text

X-Axis: miles (Q-Ratio)

Y-Axis: price (Q-Ratio)

Label: year (Nominal)

+

Mark: line

X-Axis: miles (Q-Ratio)

Y-Axis: price (Q-Ratio)

Order: year (Ordinal)

Product Pricing Solutions Explore Learn Community Search New

MIT Visualization Group Follow Share... 10

Public By Arvind Satyanarayan Edited Feb 7 Fork of Data Types, Graphical Marks, and Visual Encoding Channels

+ Add tags

Practicing Data Types, Graphical Marks, and Visual Encoding Channels with Vega-Lite

About This Reading

This interactive reading will walk you through some general background about visualization techniques and important concepts we will discuss in Wednesday's class. It includes some short exercises that take the form of code blocks that you can fill in, to help you practice these concepts.

Before you get going, the first thing you want to do is sign up for an Observable account and *fork* (make a copy of) this document, so that you have your own version that you will eventually submit. To do this, simply click the "Sign in" button at the top right of this page and select your preferred sign-in method. Once you've signed in and you're back to this document, you can select the "Fork" button, also in the top right. Every exercise you do below should be done on your own fork of this document so that your work is saved. Your final submission of this exercise will be a link to your completed fork of the reading.

Note that this notebook does include some initial snippets of JavaScript code to load a dataset. If you are unfamiliar with JavaScript, don't stress—we'll be covering these

Interactive workbook
[https://observablehq.com/
@mitvis/intro-vega-lite](https://observablehq.com/@mitvis/intro-vega-lite)

PART ONE

Data & Image Models

PART TWO

Exploratory Visual Analysis

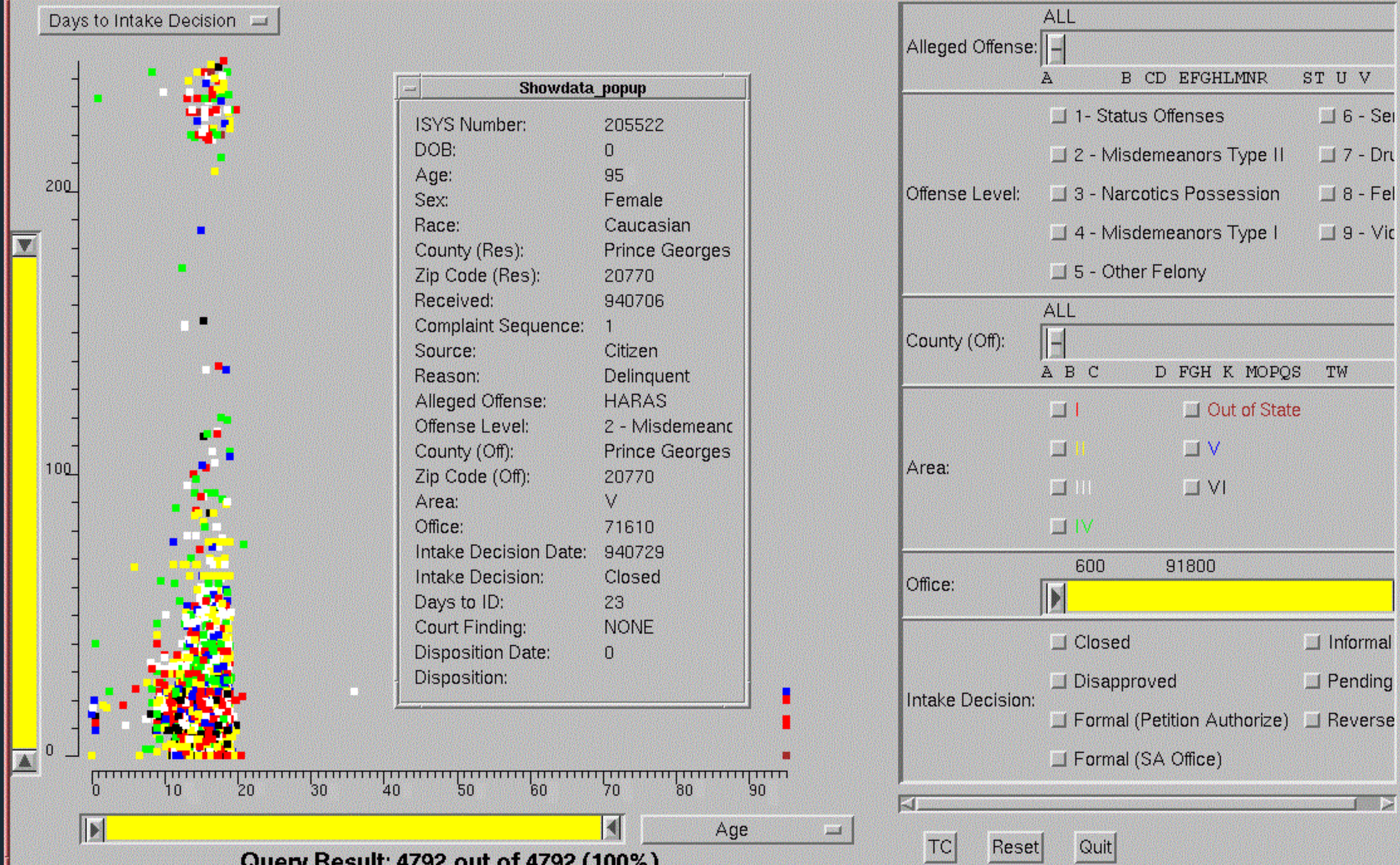
“The first sign that a visualization is good is that **it shows you a problem in your data**. Every successful visualization that I've been involved with has had this stage where you realize, **"Oh my God, this data is not what I thought it would be!"** So already, you've discovered something.”

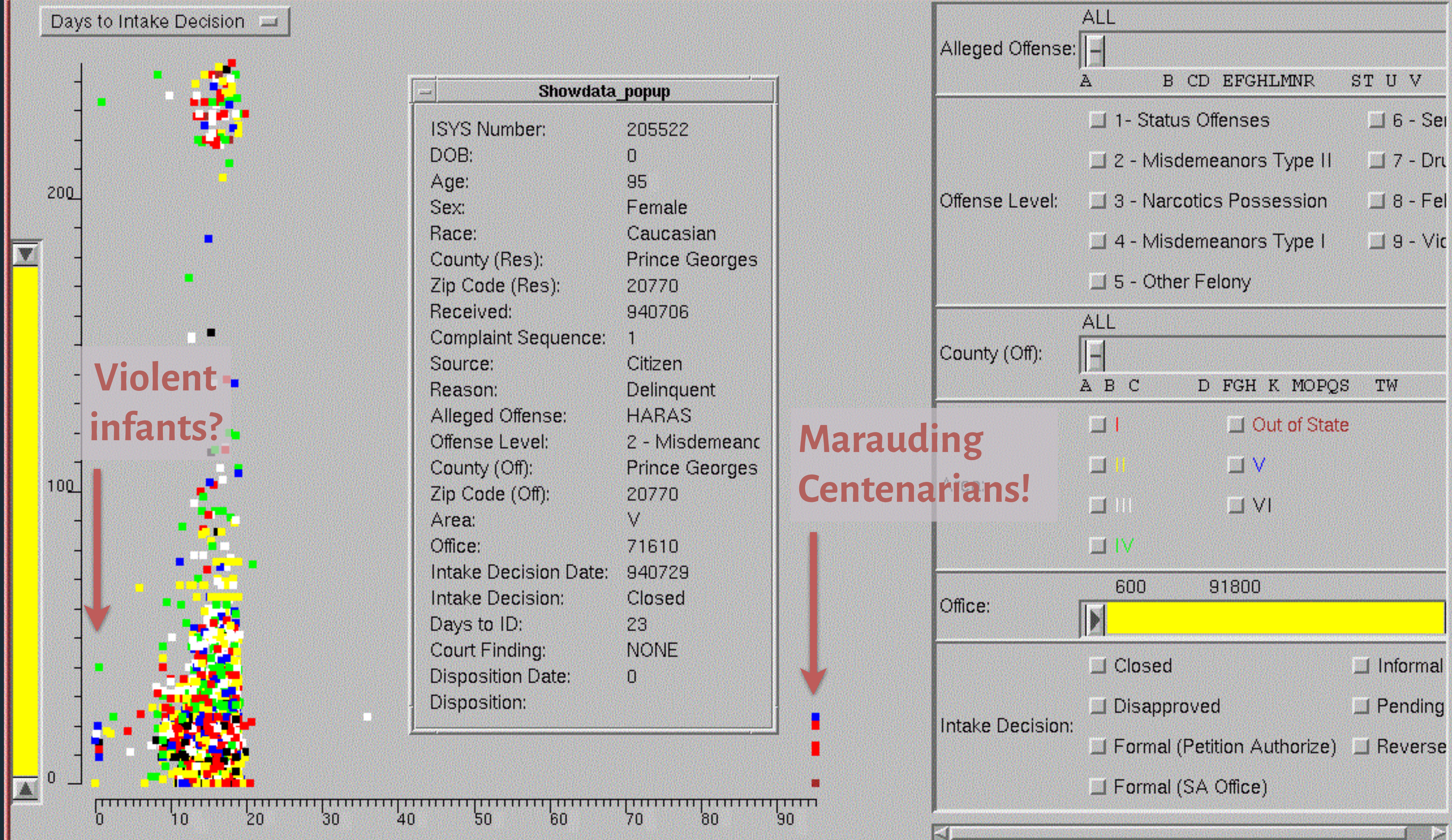
– **Martin Wattenberg**

Co-lead of Google's People + AI Initiative

ACM Queue, Mar 2010







Violent infants?

Marauding Centenarians!

Showdata_popup	
ISYS Number:	205522
DOB:	0
Age:	95
Sex:	Female
Race:	Caucasian
County (Res):	Prince Georges
Zip Code (Res):	20770
Received:	940706
Complaint Sequence:	1
Source:	Citizen
Reason:	Delinquent
Alleged Offense:	HARAS
Offense Level:	2 - Misdemeanor
County (Off):	Prince Georges
Zip Code (Off):	20770
Area:	V
Office:	71610
Intake Decision Date:	940729
Intake Decision:	Closed
Days to ID:	23
Court Finding:	NONE
Disposition Date:	0
Disposition:	

Alleged Offense: [ALL] [A] [B] [CD] [EFGHLMNR] [ST] [U] [V]

Offense Level: 1 - Status Offenses 6 - Sex
 2 - Misdemeanors Type II 7 - Dru
 3 - Narcotics Possession 8 - Fel
 4 - Misdemeanors Type I 9 - Vic
 5 - Other Felony

County (Off): [ALL] [A] [B] [C] [D] [FGH] [K] [MOPQS] [TW]

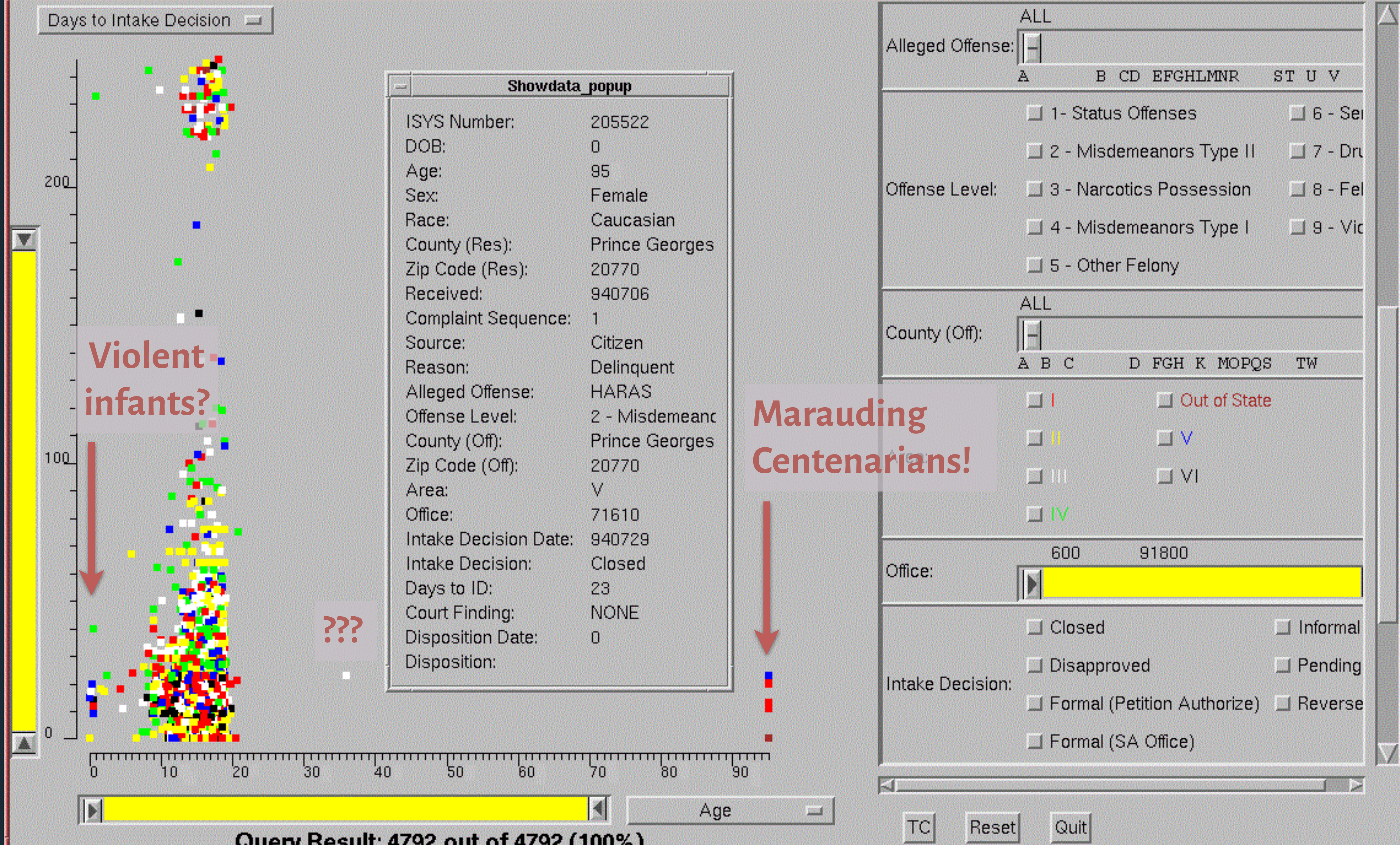
I Out of State
 II V
 III VI
 IV

Office: 600 91800

Intake Decision: Closed Informal
 Disapproved Pending
 Formal (Petition Authorize) Reverse
 Formal (SA Office)

Query Result: 4792 out of 4792 (100%)

TC Reset Quit



Exploratory Visual Analysis

Process

1. Construct graphics to address questions.
2. Inspect "answer" and ask new questions.
3. Iterate...

→ Trends



→ Outliers



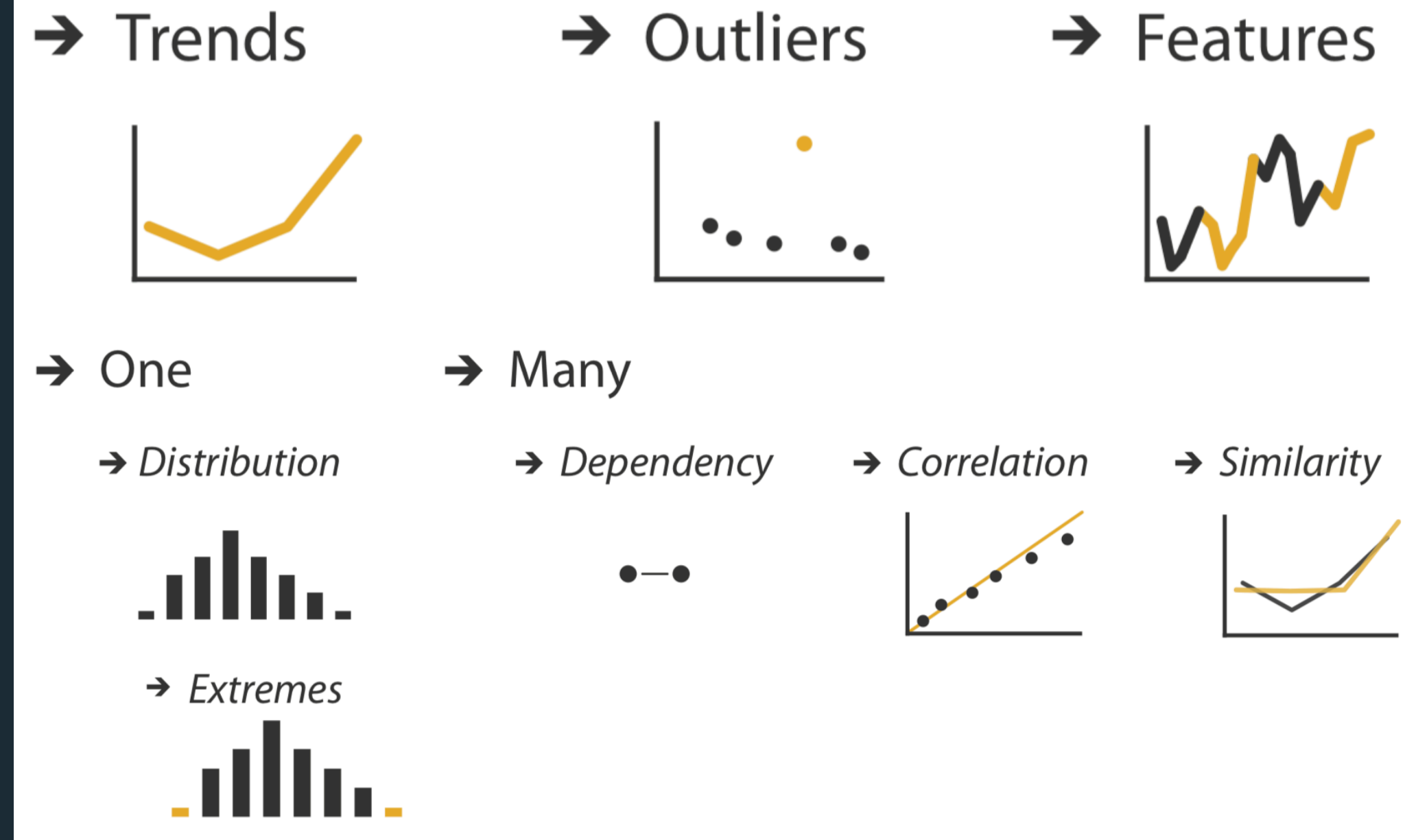
→ Features



Exploratory Visual Analysis

Process

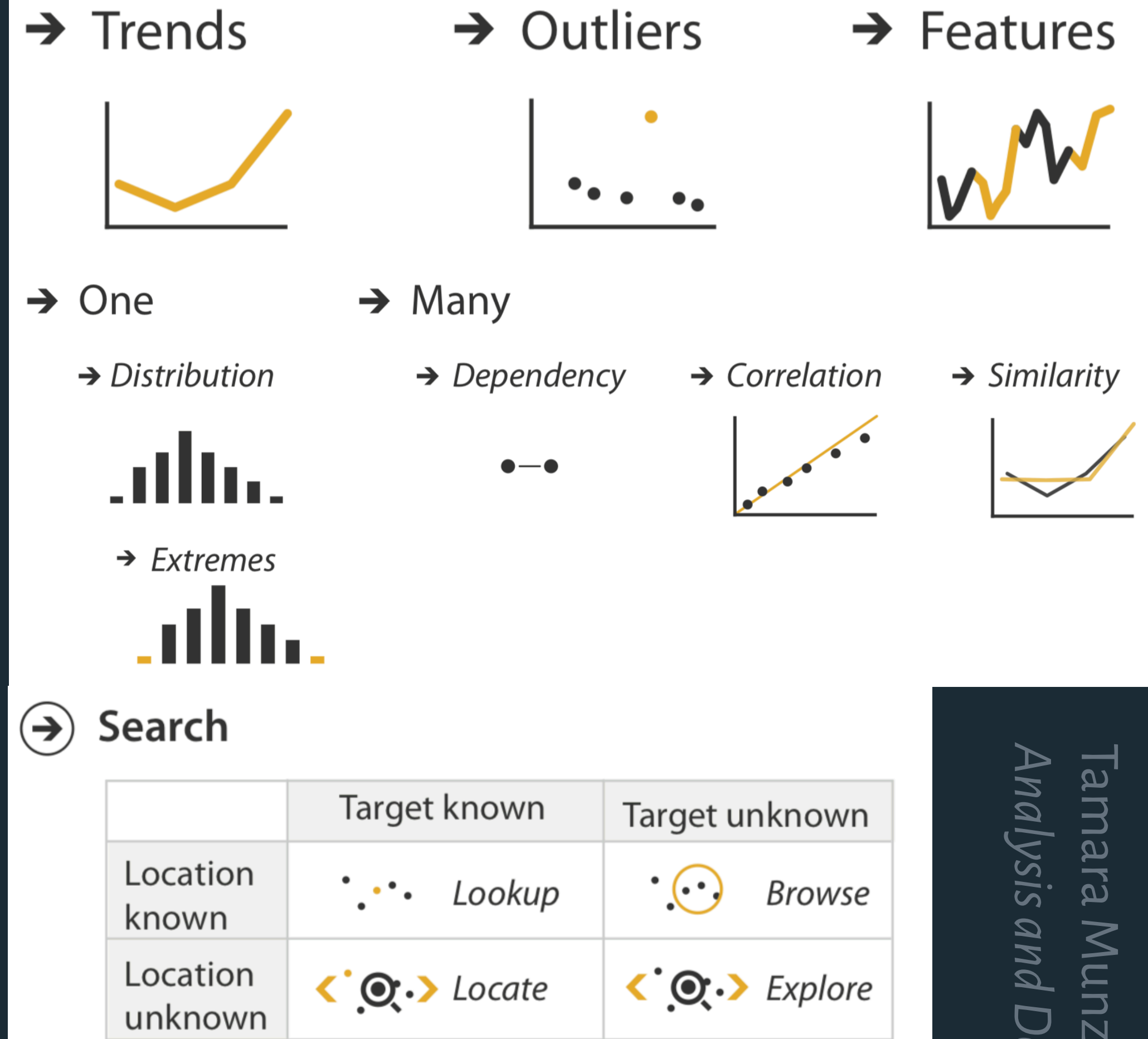
1. Construct graphics to address questions.
2. Inspect "answer" and ask new questions.
3. Iterate...



Exploratory Visual Analysis

Process

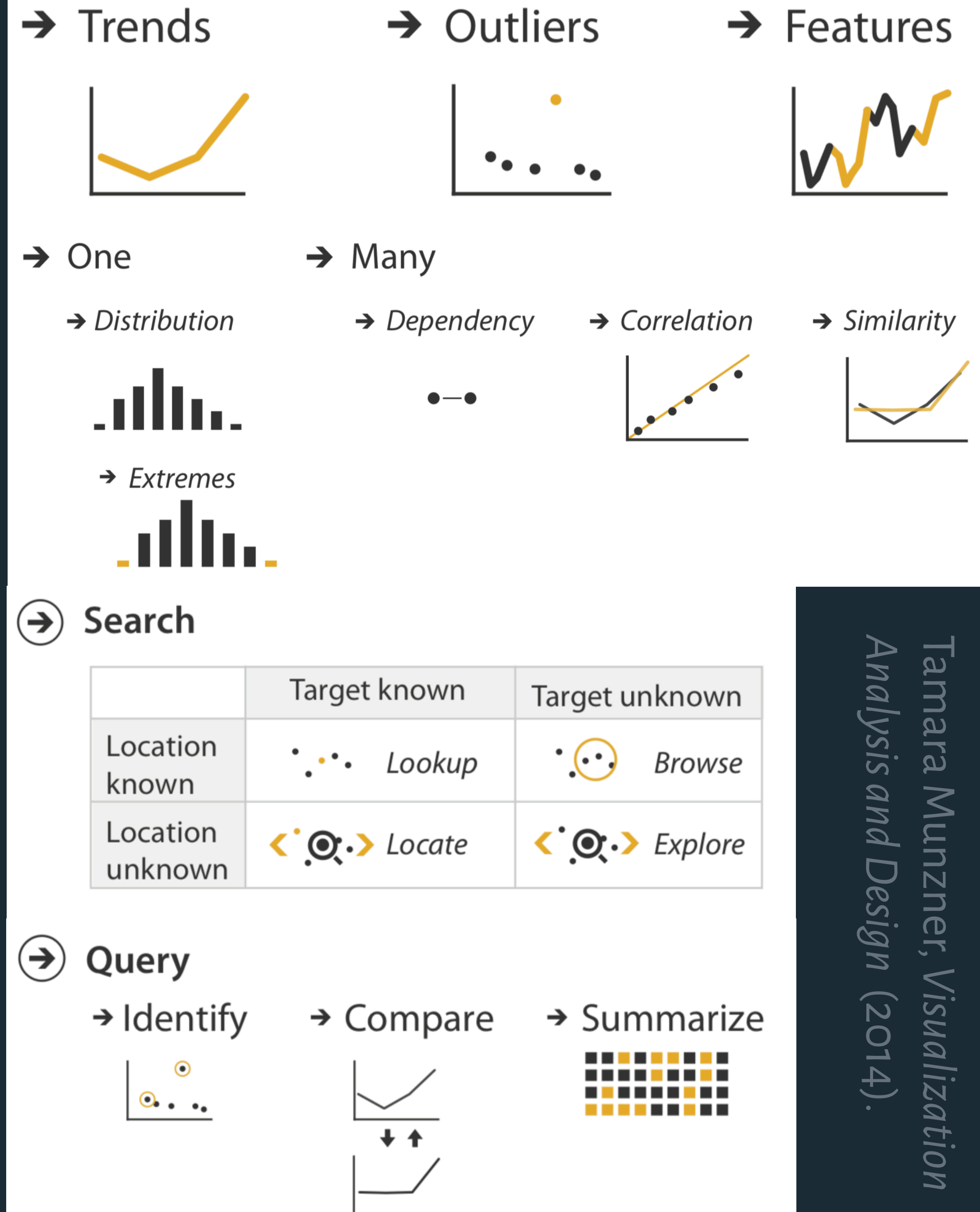
1. Construct graphics to address questions.
2. Inspect "answer" and ask new questions.
3. Iterate...



Exploratory Visual Analysis

Process

1. Construct graphics to address questions.
2. Inspect "answer" and ask new questions.
3. Iterate...





Graph Viewer

Roll-up by:

All

Visualization:

Node-Link

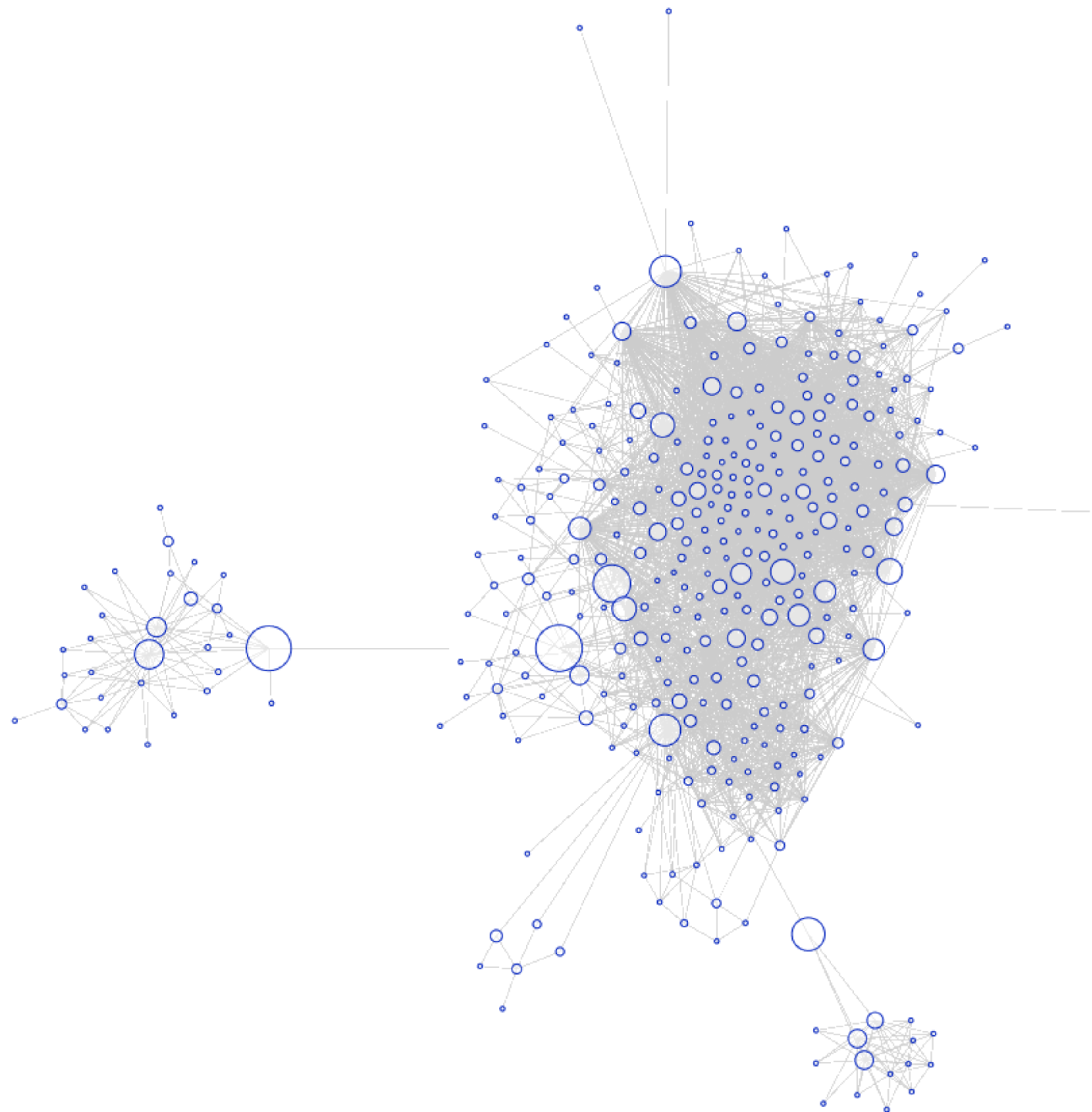
Sort by:

None

Edge centrality filters:

Two horizontal sliders for edge centrality filtering, both currently set to the minimum value.

- Images
- Animate





Graph Viewer

Roll-up by:

All

Visualization:

Matrix

Sort by:

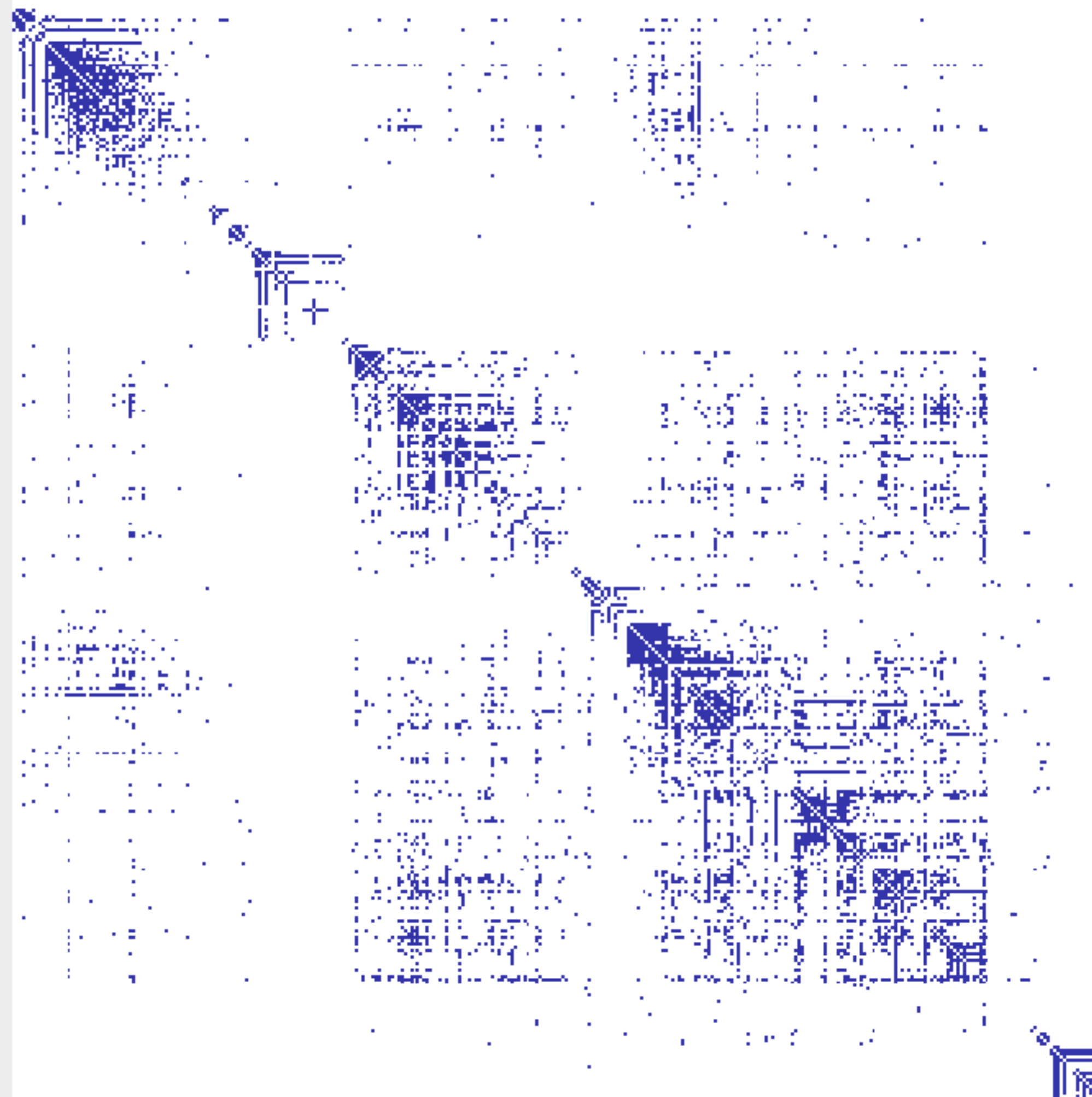
Linkage

Edge centrality filters:

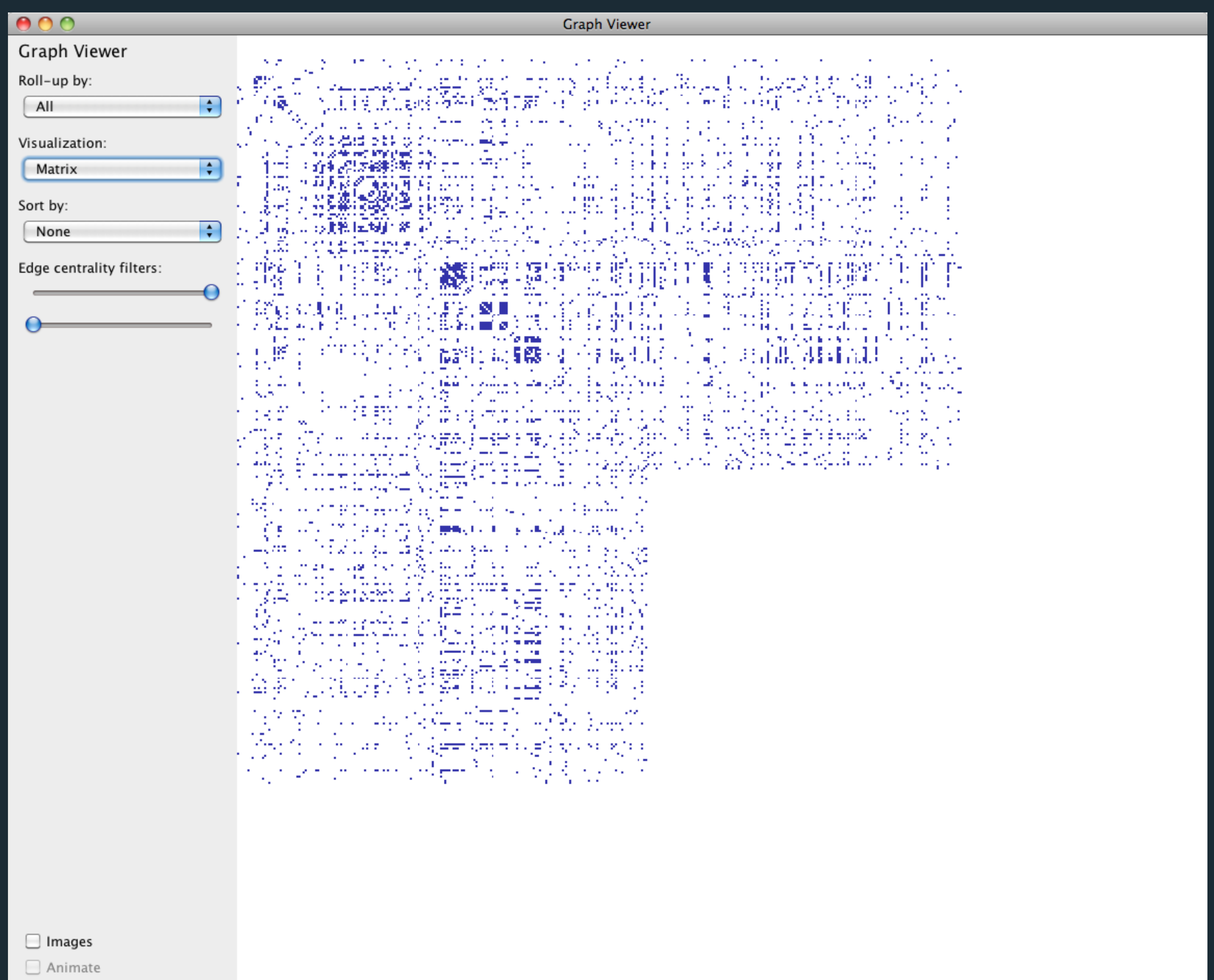
Two horizontal sliders for edge centrality filters, both currently set to the minimum value.

Images

Animate



Missing Values



Exploratory Visual Analysis




Process

1. Construct graphics to address questions.
2. Inspect "answer" and ask new questions.
3. Iterate...

Lessons



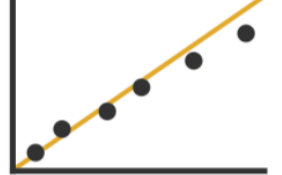

- ✓ Check **data quality** and your **assumptions**.
- ✓ Start with **univariate summaries**, then consider **relationships between variables**.

→ Trends → Outliers → Features






→ One → Many





→ Distribution → Dependency → Correlation → Similarity

→ Extremes






→ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

→ Query

→ Identify → Compare → Summarize

Analysis Example: Antibiotic Effectiveness

Analysis Example: Antibiotic Effectiveness

Collected prior to 1951

Genus of Bacteria String (N)

Species of Bacteria String (N)

Antibiotic Applied String (N)

Gram-Staining? Pos / Neg (N)

Min. Inhibitory Con. (g) Number (Q)

Analysis Example: Antibiotic Effectiveness

Collected prior to 1951

Genus of Bacteria

String (N)

Species of Bacteria

String (N)

Antibiotic Applied

String (N)

Gram-Staining?

Pos / Neg (N)

Min. Inhibitory Con. (g)

Number (Q)

Table 1—Burtin's Data

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
<i>Aerobacter aerogenes</i>	870	1	1.6	negative
<i>Brucella abortus</i>	1	2	0.02	negative
<i>Brucella anthracis</i>	0.001	0.01	0.007	positive
<i>Diplococcus pneumoniae</i>	0.005	11	10	positive
<i>Escherichia coli</i>	100	0.4	0.1	negative
<i>Klebsiella pneumoniae</i>	850	1.2	1	negative
<i>Mycobacterium tuberculosis</i>	800	5	2	negative
<i>Proteus vulgaris</i>	3	0.1	0.1	negative
<i>Pseudomonas aeruginosa</i>	850	2	0.4	negative
<i>Salmonella (Eberthella) typhosa</i>	1	0.4	0.008	negative
<i>Salmonella schottmuelleri</i>	10	0.8	0.09	negative
<i>Staphylococcus albus</i>	0.007	0.1	0.001	positive
<i>Staphylococcus aureus</i>	0.03	0.03	0.001	positive
<i>Streptococcus fecalis</i>	1	1	0.1	positive
<i>Streptococcus hemolyticus</i>	0.001	14	10	positive
<i>Streptococcus viridans</i>	0.005	10	40	positive

What questions might we ask?

Collected prior to 1951

Genus of Bacteria

String (N)

Species of Bacteria

String (N)

Antibiotic Applied

String (N)

Gram-Staining?

Pos / Neg (N)

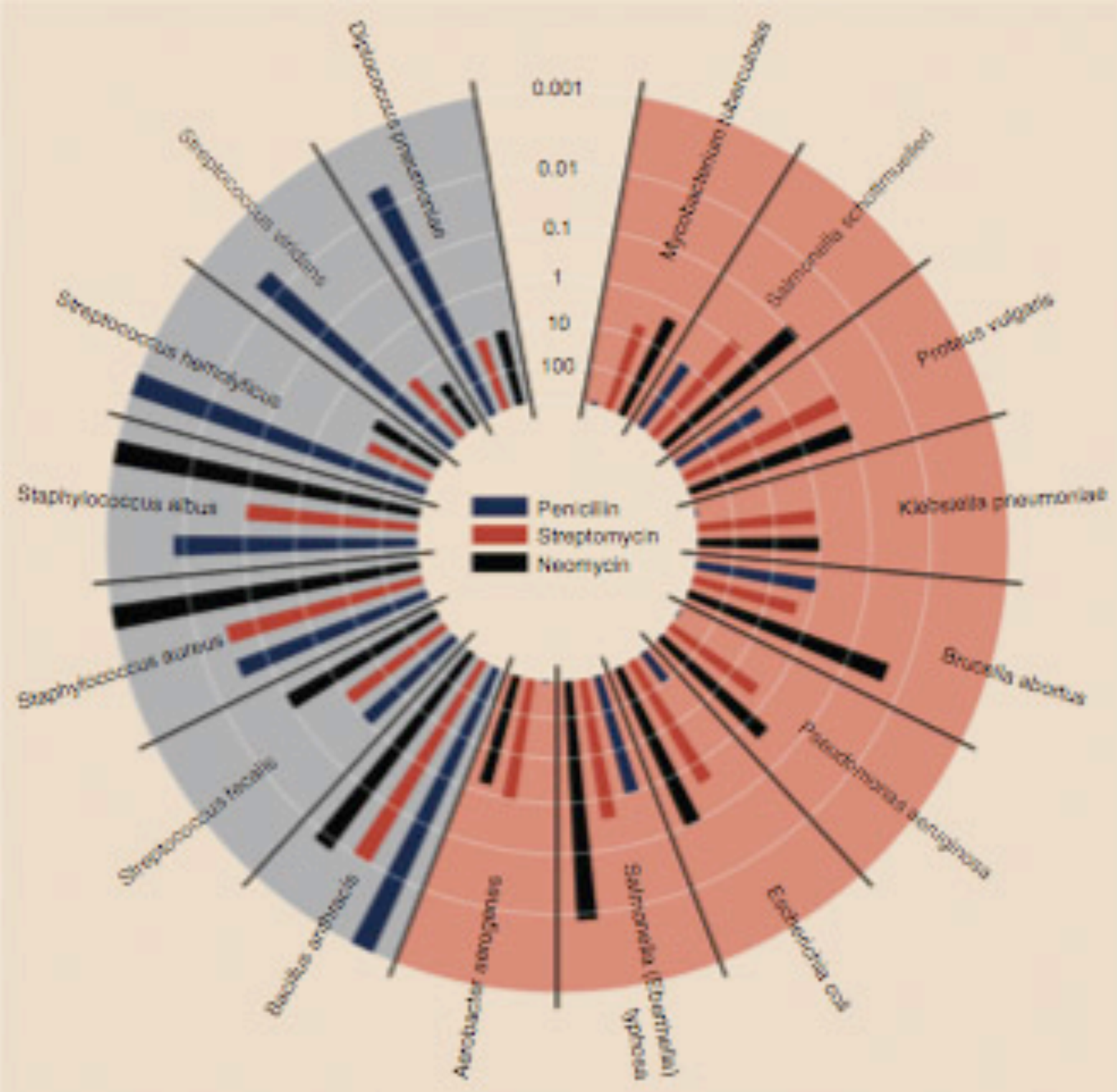
Min. Inhibitory Con. (g)

Number (Q)

Table 1—Burtin's Data

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
<i>Aerobacter aerogenes</i>	870	1	1.6	negative
<i>Brucella abortus</i>	1	2	0.02	negative
<i>Brucella anthracis</i>	0.001	0.01	0.007	positive
<i>Diplococcus pneumoniae</i>	0.005	11	10	positive
<i>Escherichia coli</i>	100	0.4	0.1	negative
<i>Klebsiella pneumoniae</i>	850	1.2	1	negative
<i>Mycobacterium tuberculosis</i>	800	5	2	negative
<i>Proteus vulgaris</i>	3	0.1	0.1	negative
<i>Pseudomonas aeruginosa</i>	850	2	0.4	negative
<i>Salmonella (Eberthella) typhosa</i>	1	0.4	0.008	negative
<i>Salmonella schottmuelleri</i>	10	0.8	0.09	negative
<i>Staphylococcus albus</i>	0.007	0.1	0.001	positive
<i>Staphylococcus aureus</i>	0.03	0.03	0.001	positive
<i>Streptococcus fecalis</i>	1	1	0.1	positive
<i>Streptococcus hemolyticus</i>	0.001	14	10	positive
<i>Streptococcus viridans</i>	0.005	10	40	positive

How do the drugs compare?



Bacteria	Penicillin	Antibiotic Streptomycin	Neomycin	Gram stain
<i>Aerobacter aerogenes</i>	870	1	1.6	-
<i>Brucella abortus</i>	1	2	0.02	-
<i>Bacillus anthracis</i>	0.001	0.01	0.007	+
<i>Diplococcus pneumoniae</i>	0.005	11	10	+
<i>Escherichia coli</i>	100	0.4	0.1	-
<i>Klebsiella pneumoniae</i>	850	1.2	1	-
<i>Mycobacterium tuberculosis</i>	800	5	2	-
<i>Proteus vulgaris</i>	3	0.1	0.1	-
<i>Pseudomonas aeruginosa</i>	850	2	0.4	-
<i>Salmonella (Eberthella) typhosa</i>	1	0.4	0.008	-
<i>Salmonella schottmuelleri</i>	10	0.8	0.09	-
<i>Staphylococcus albus</i>	0.007	0.1	0.001	+
<i>Staphylococcus aureus</i>	0.03	0.03	0.001	+
<i>Streptococcus fecalis</i>	1	1	0.1	+
<i>Streptococcus hemolyticus</i>	0.001	14	10	+
<i>Streptococcus viridans</i>	0.005	10	40	+

Encodings

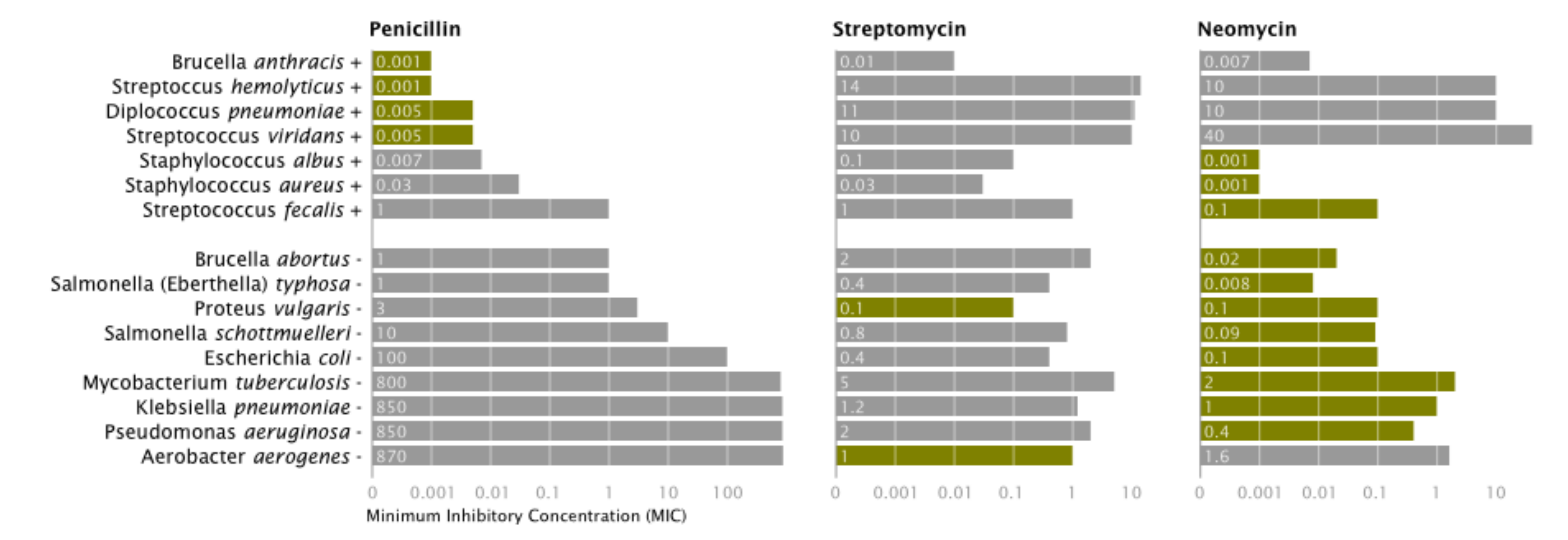
Radius: $1 / \log(\text{MIC})$

Bar Color: Antibiotic

Background Color: Gram Staining

Original graphic by Will Burtin, 1951.

How do the drugs compare?



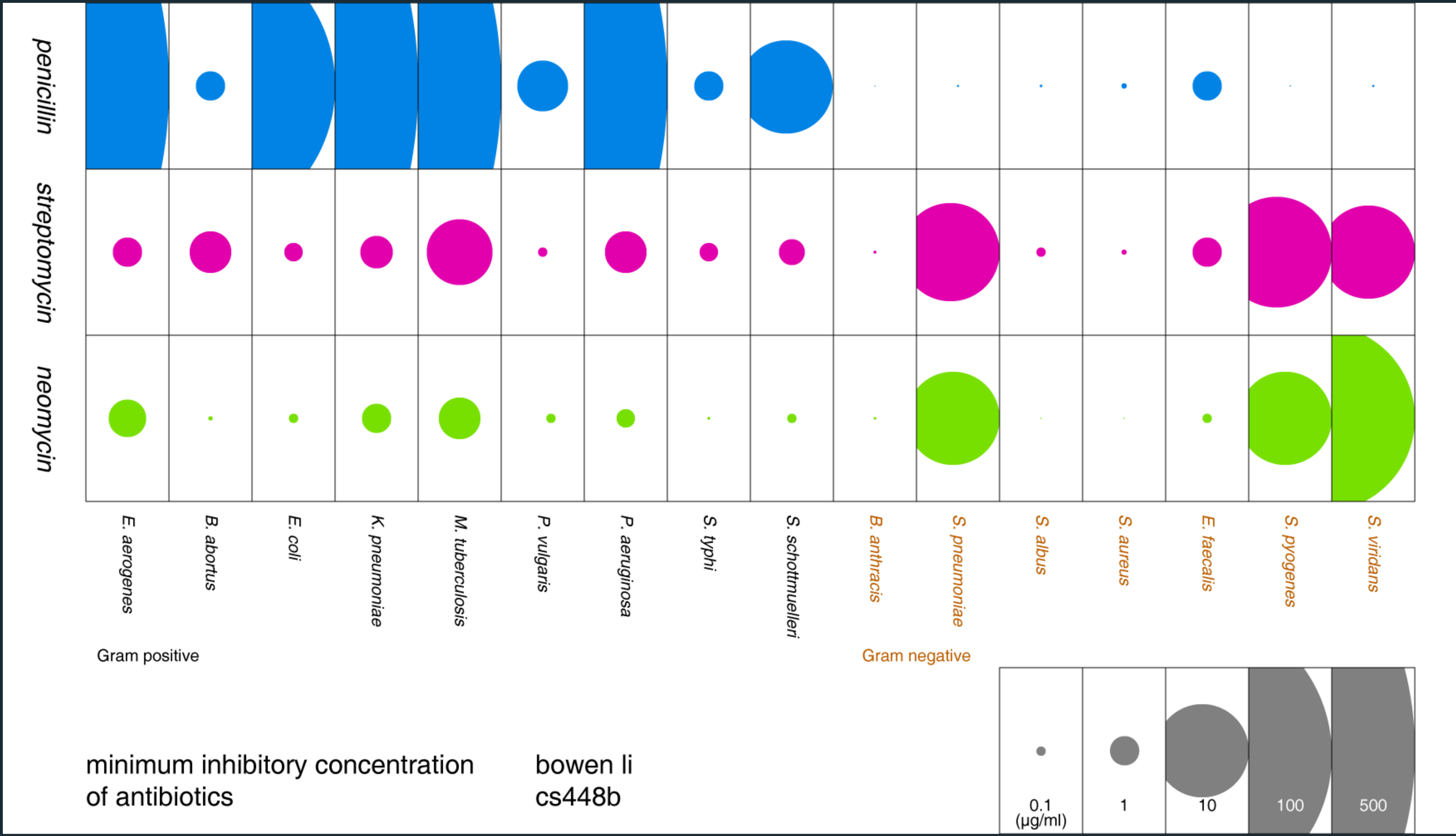
X-Axis: Antibiotic | log(MIC)

Y-Axis: Gram-Staining | Species

Color: Most Effective?

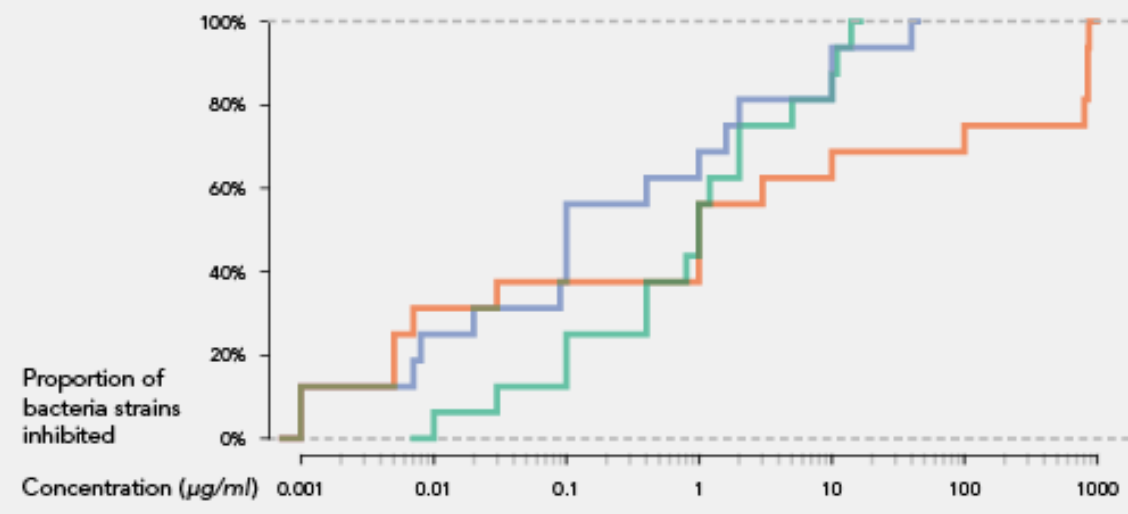
Mike Bostock, *Stanford CS448b* (Winter 2009).

How do the drugs compare?



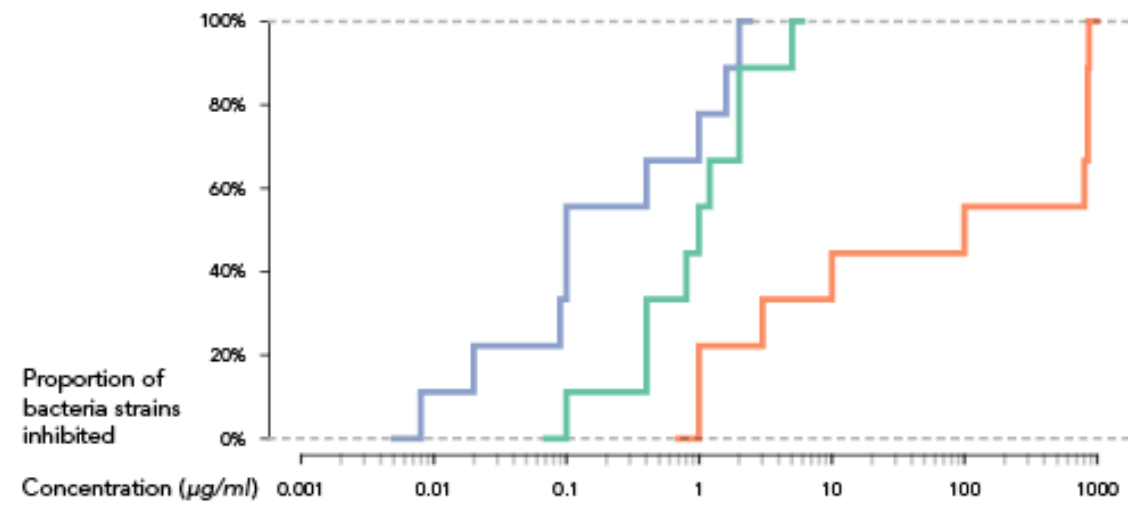
Bowen Li, Stanford CS448b (Fall 2009).

All bacteria



Streptomycin and Neomycin are more efficient broad-spectrum antibiotics than Penicillin.

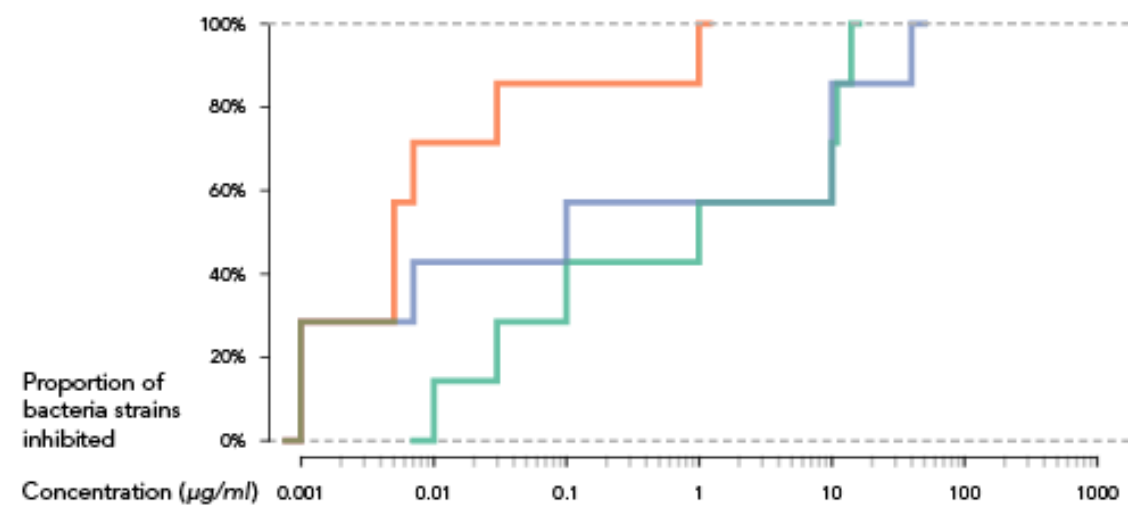
Gram-negative bacteria only



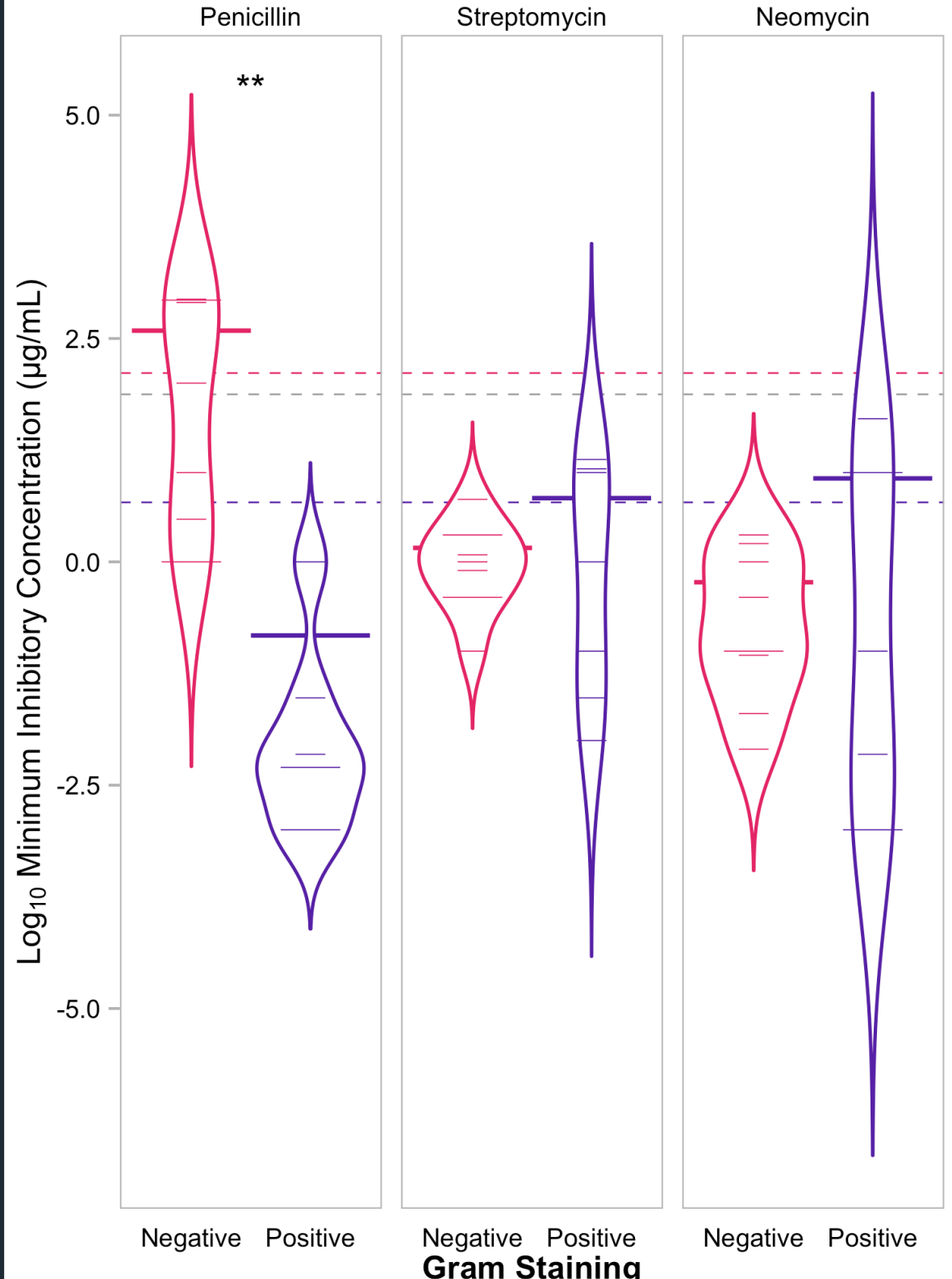
Neomycin and Streptomycin are more efficient against gram-negative bacteria, so can be used at a lower dosage here than above.

Gram staining quickly identifies bacteria as Gram-negative or Gram-positive, which can be used to find a more efficient antibiotic and dosage.

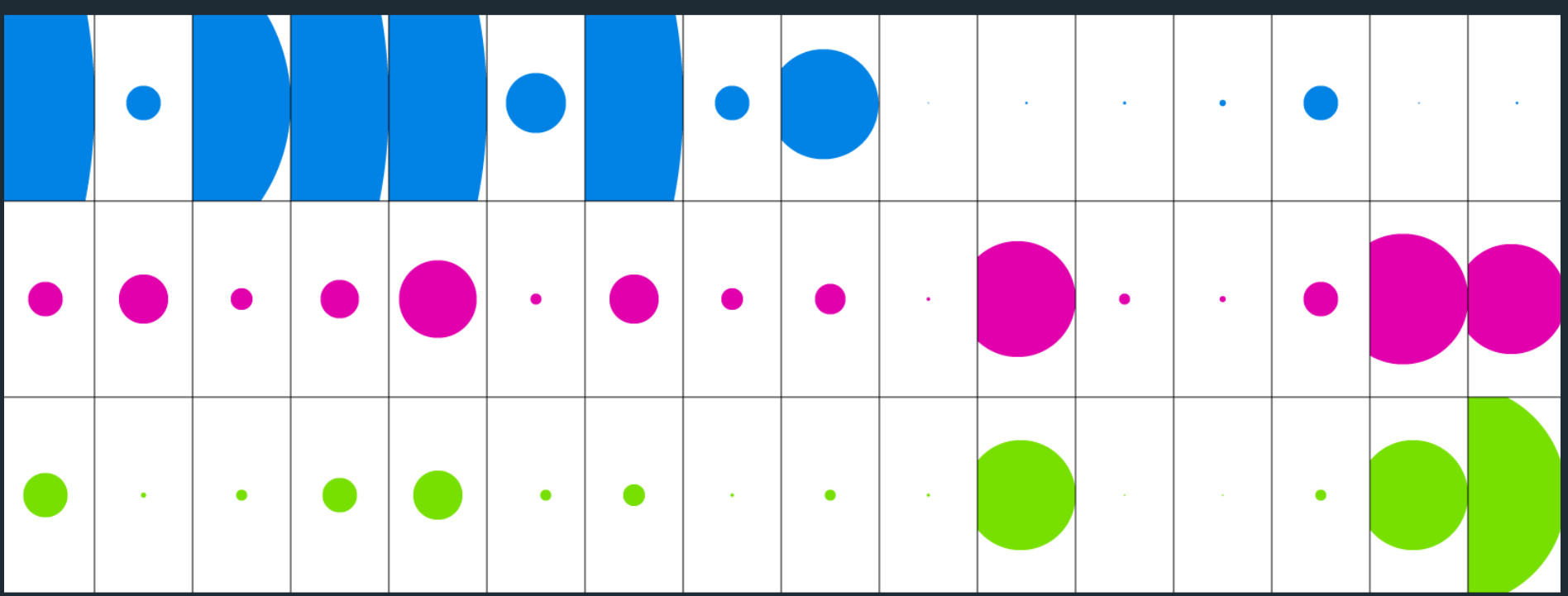
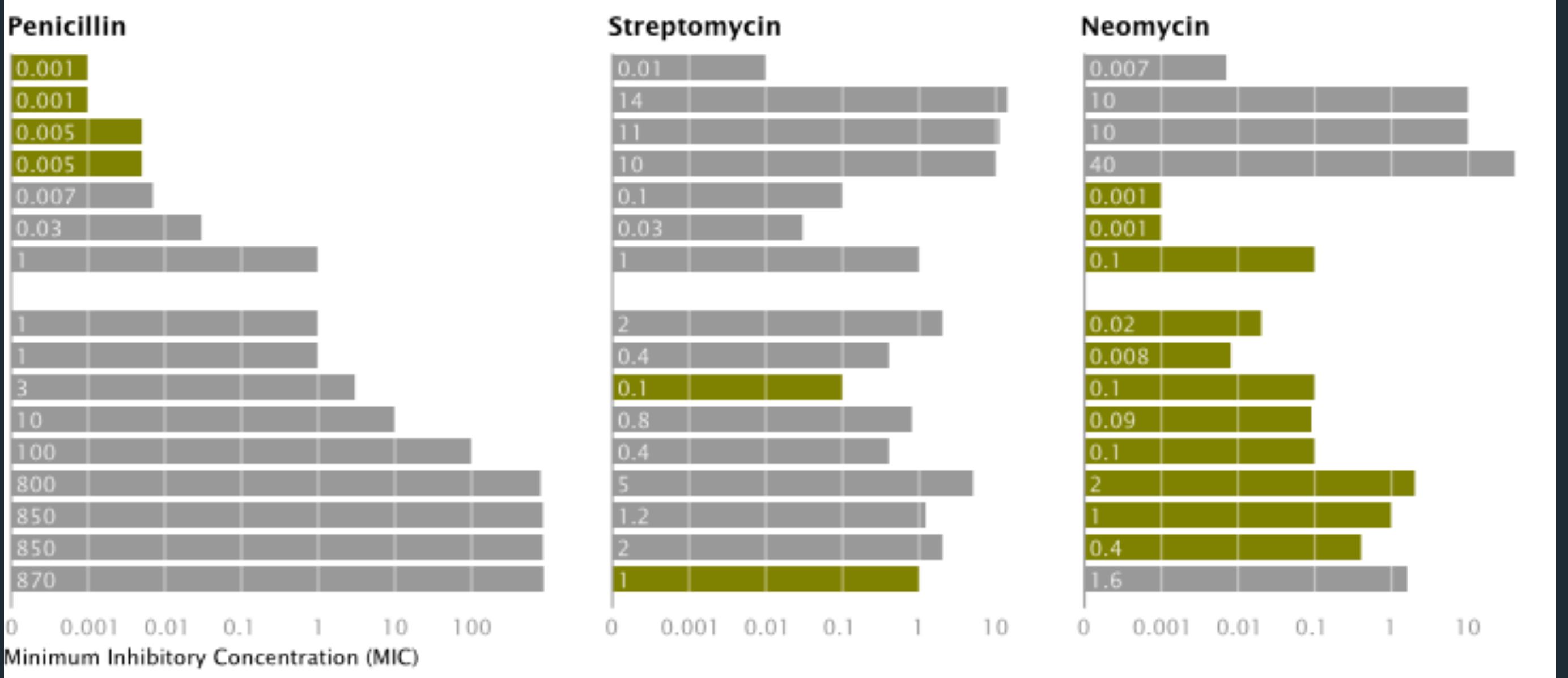
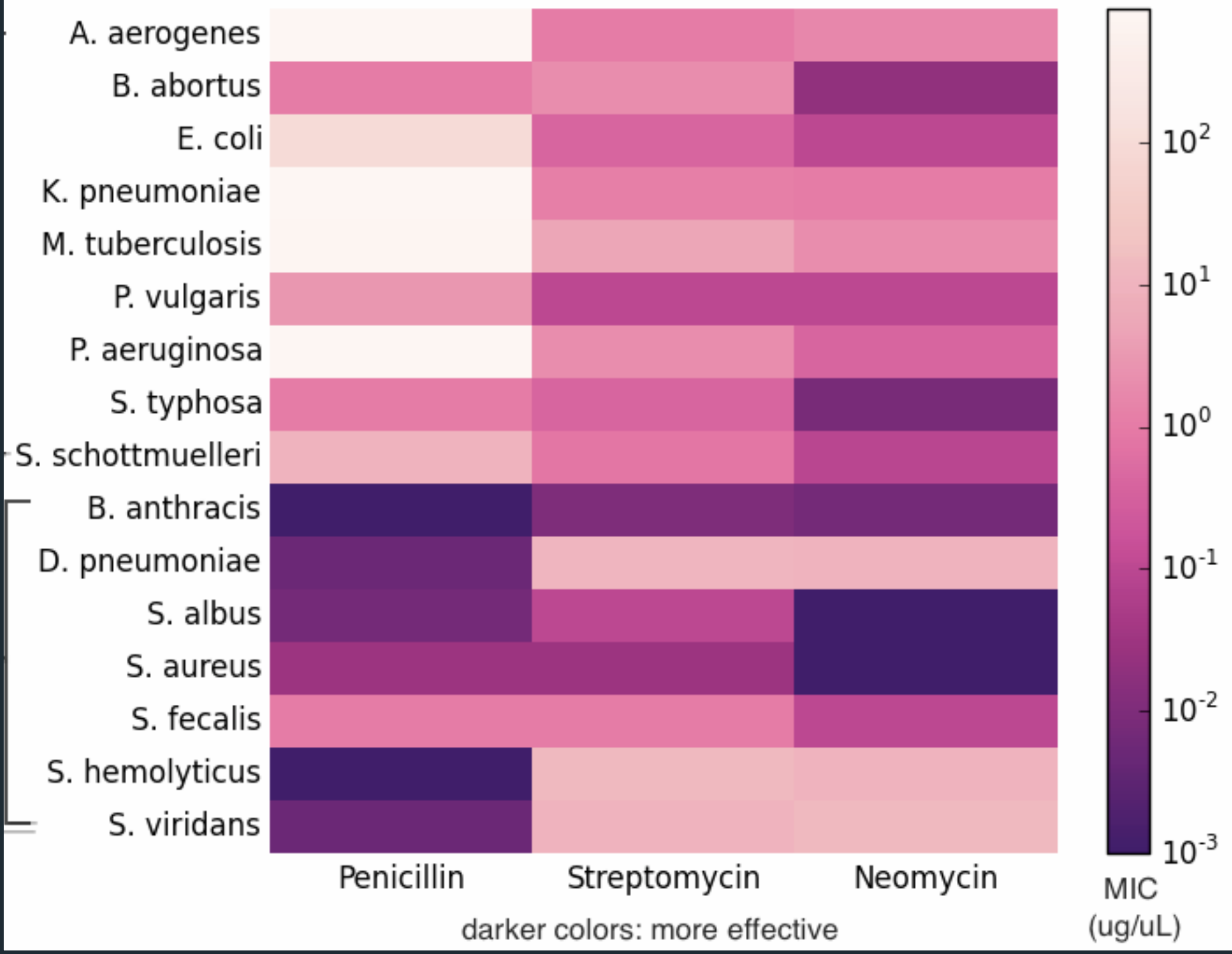
Gram-positive bacteria only

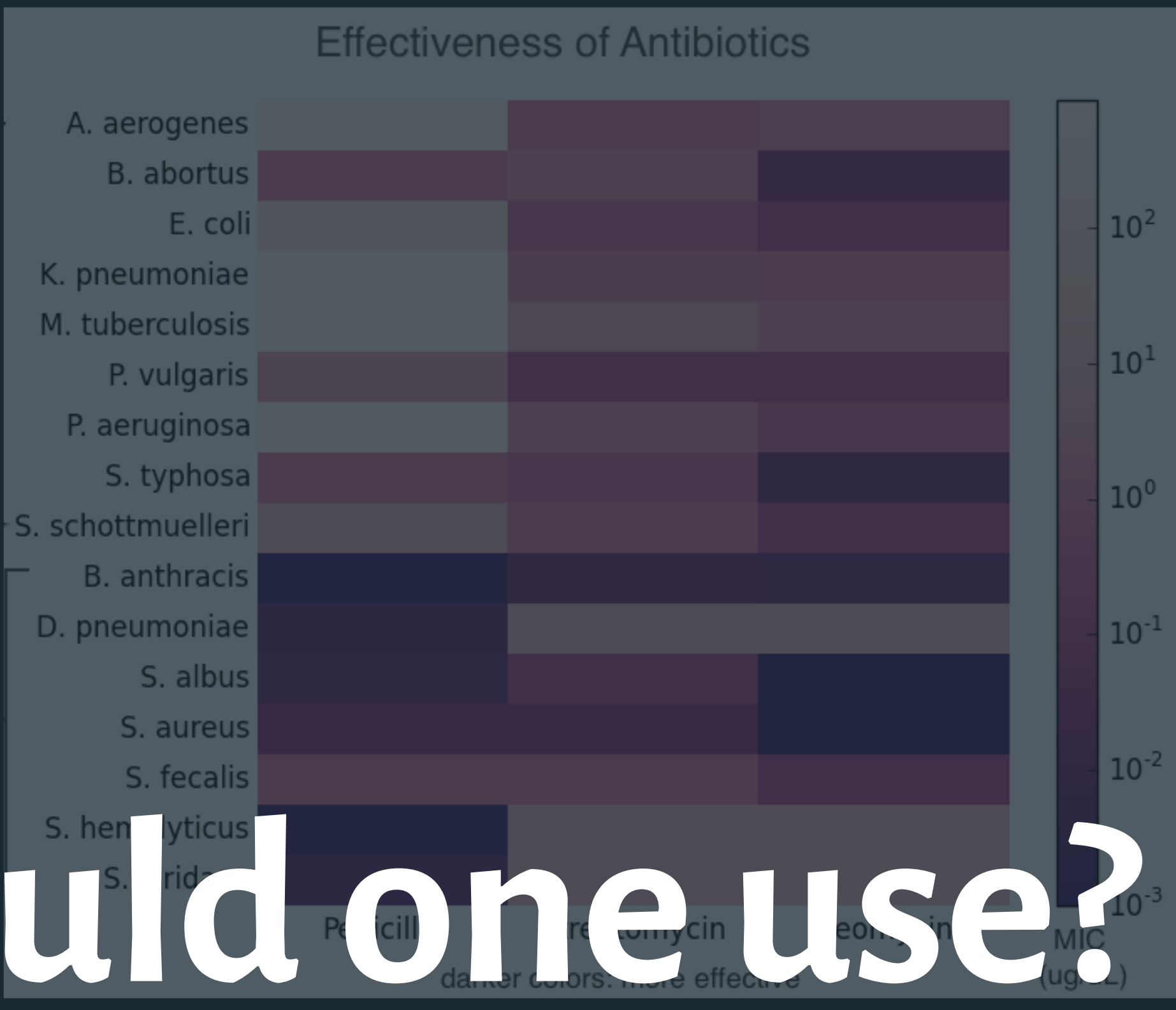
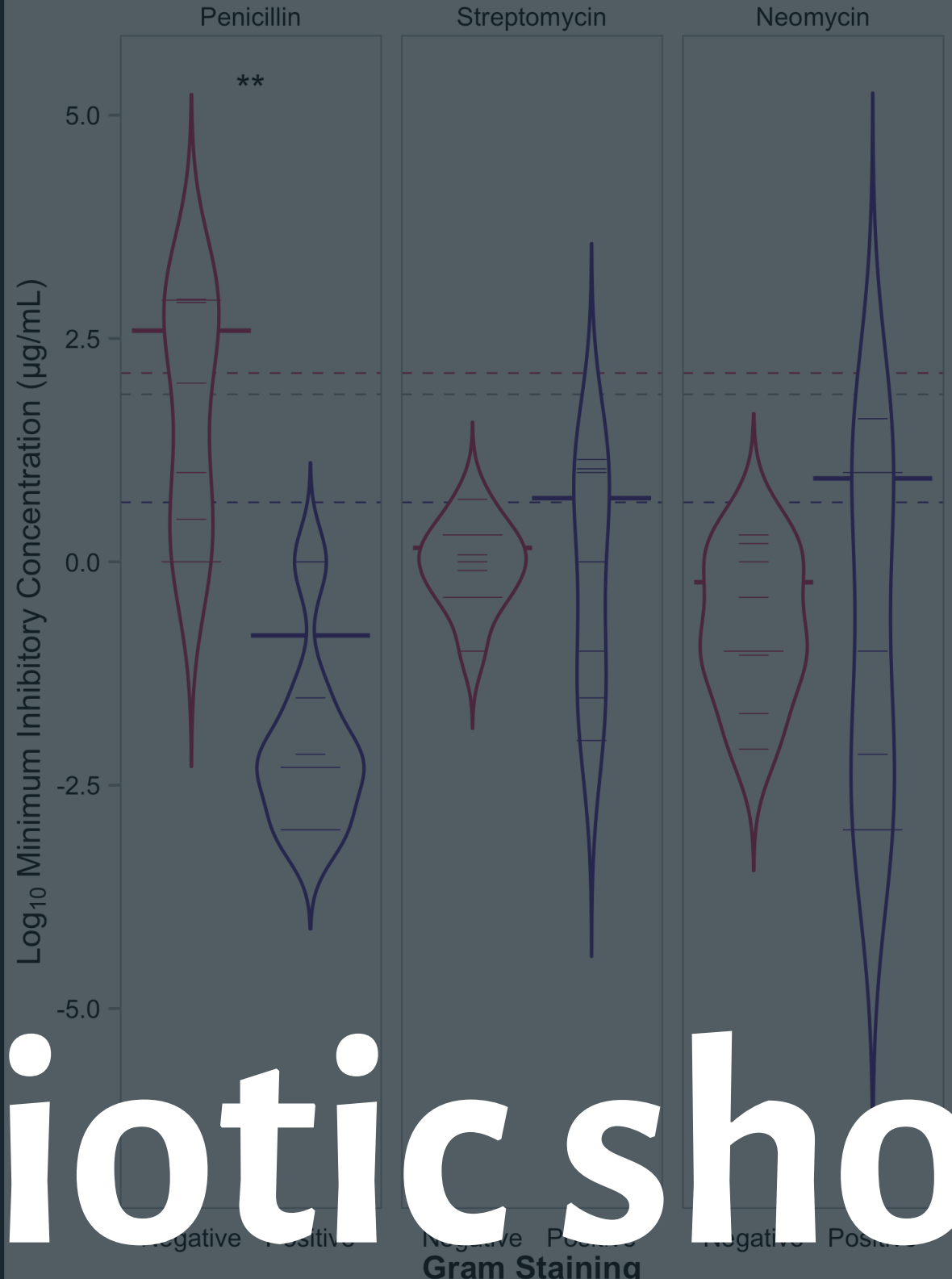
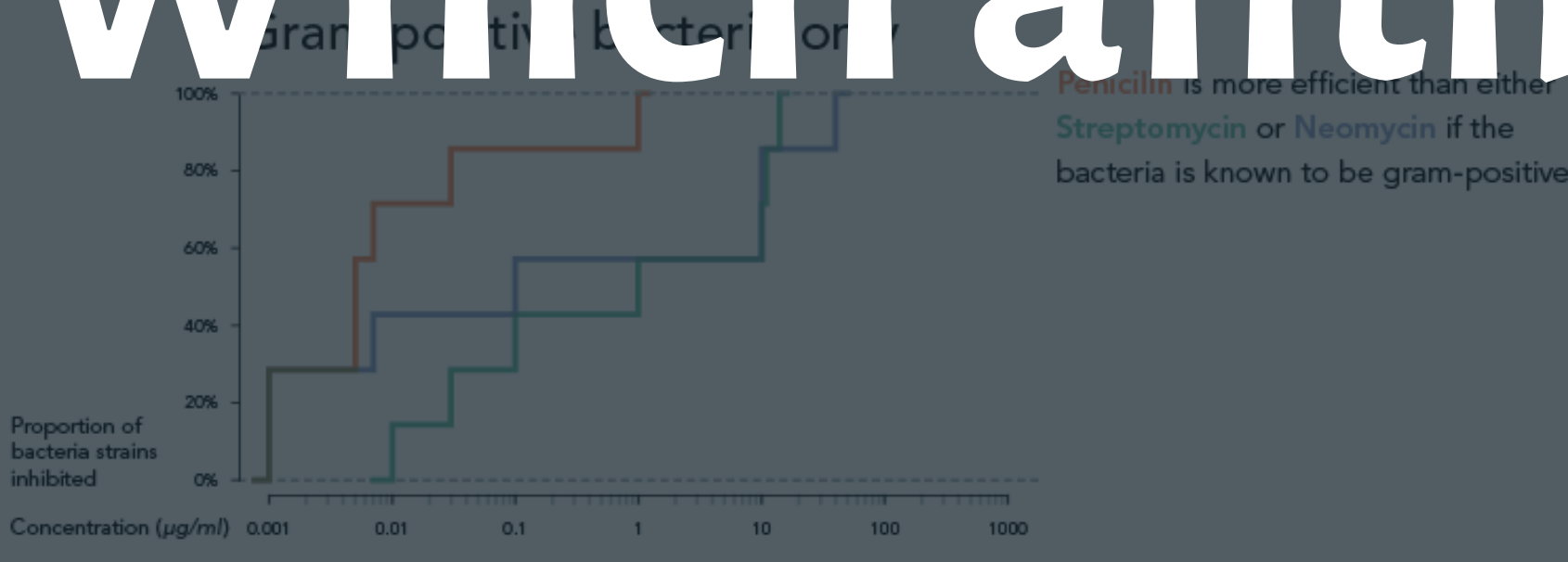
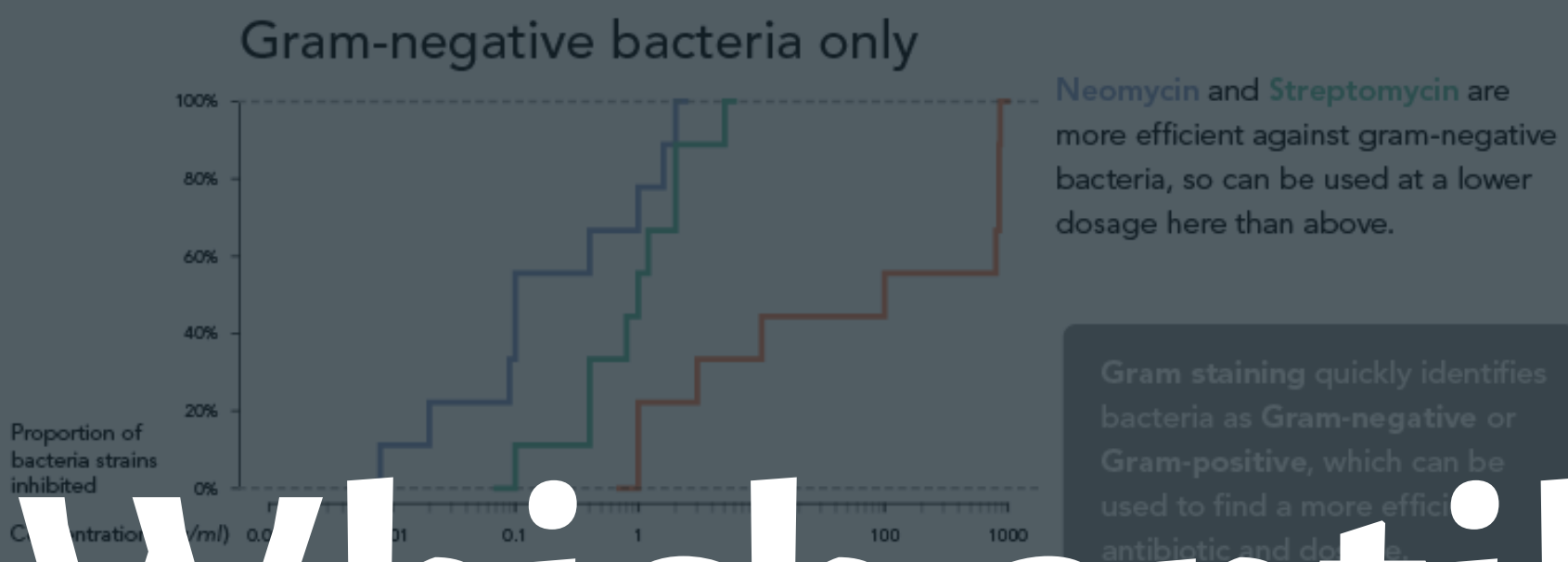
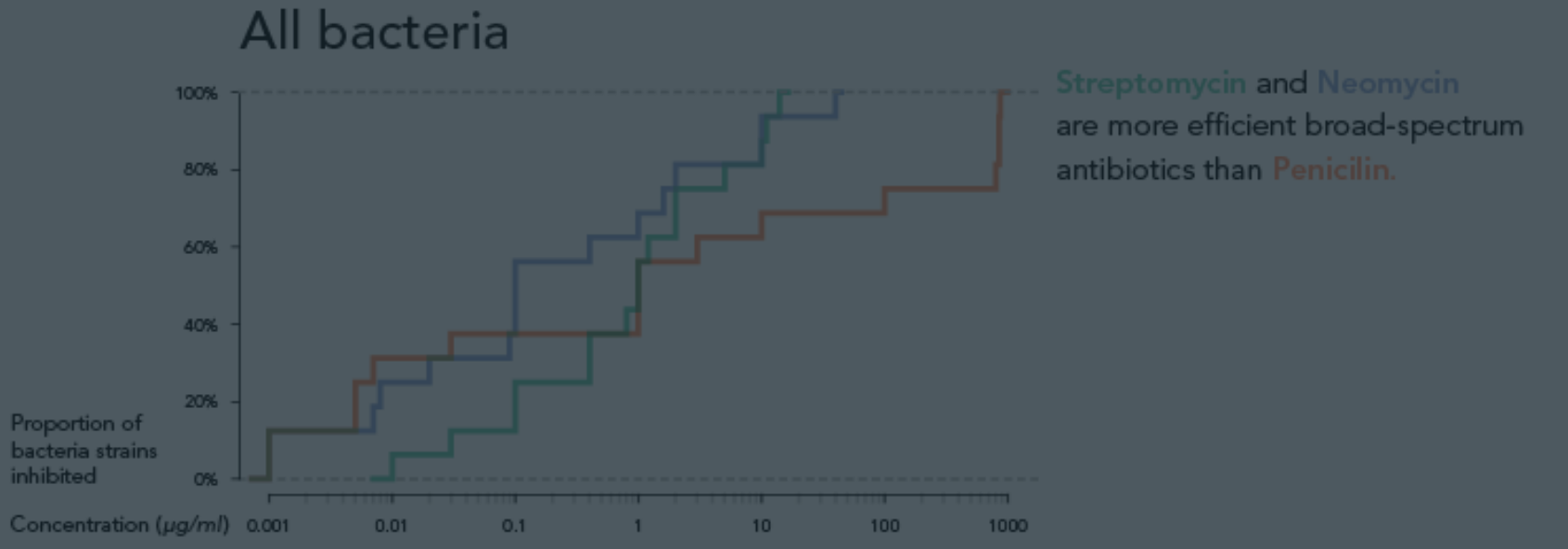


Penicillin is more efficient than either Streptomycin or Neomycin if the bacteria is known to be gram-positive.

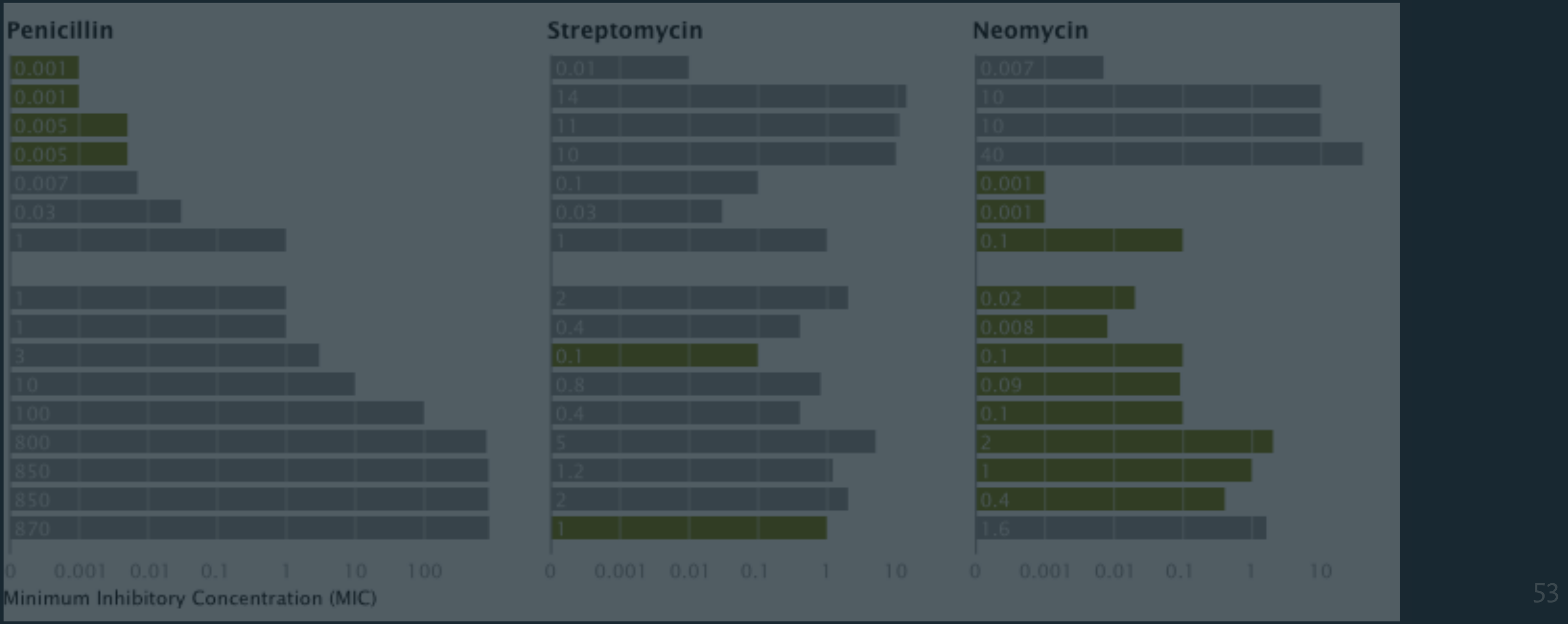
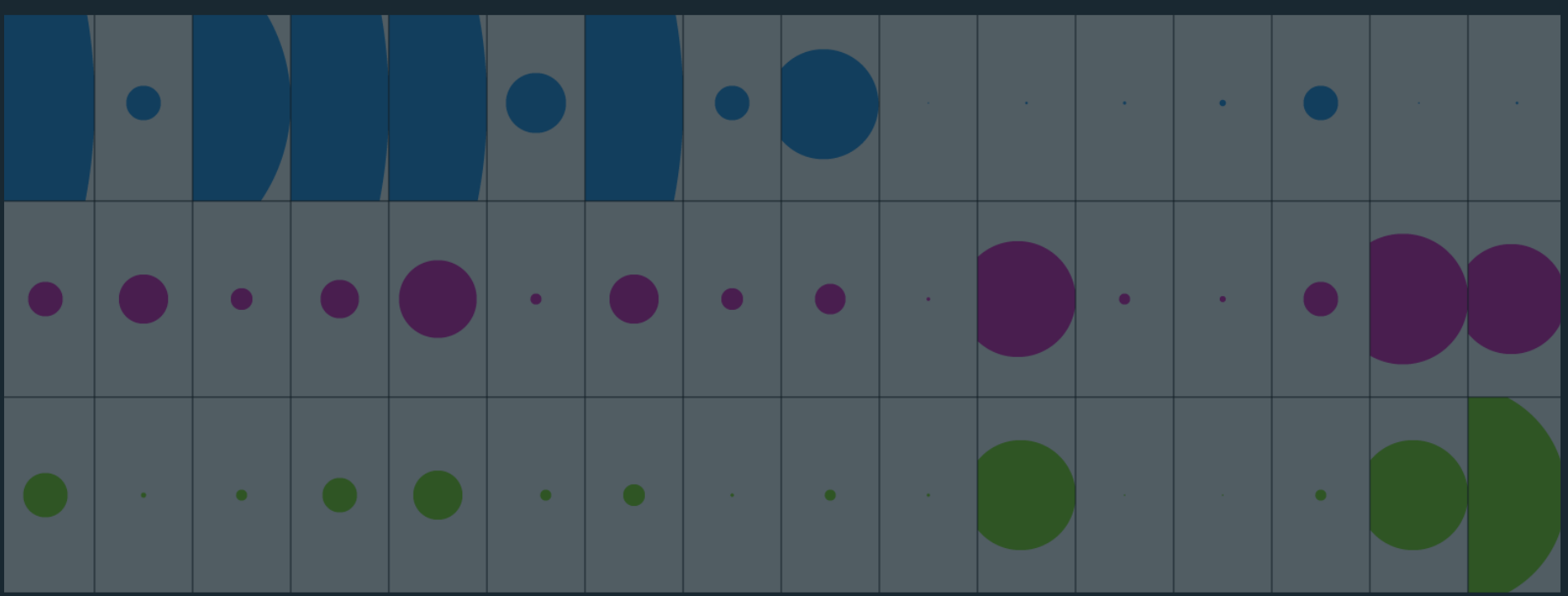


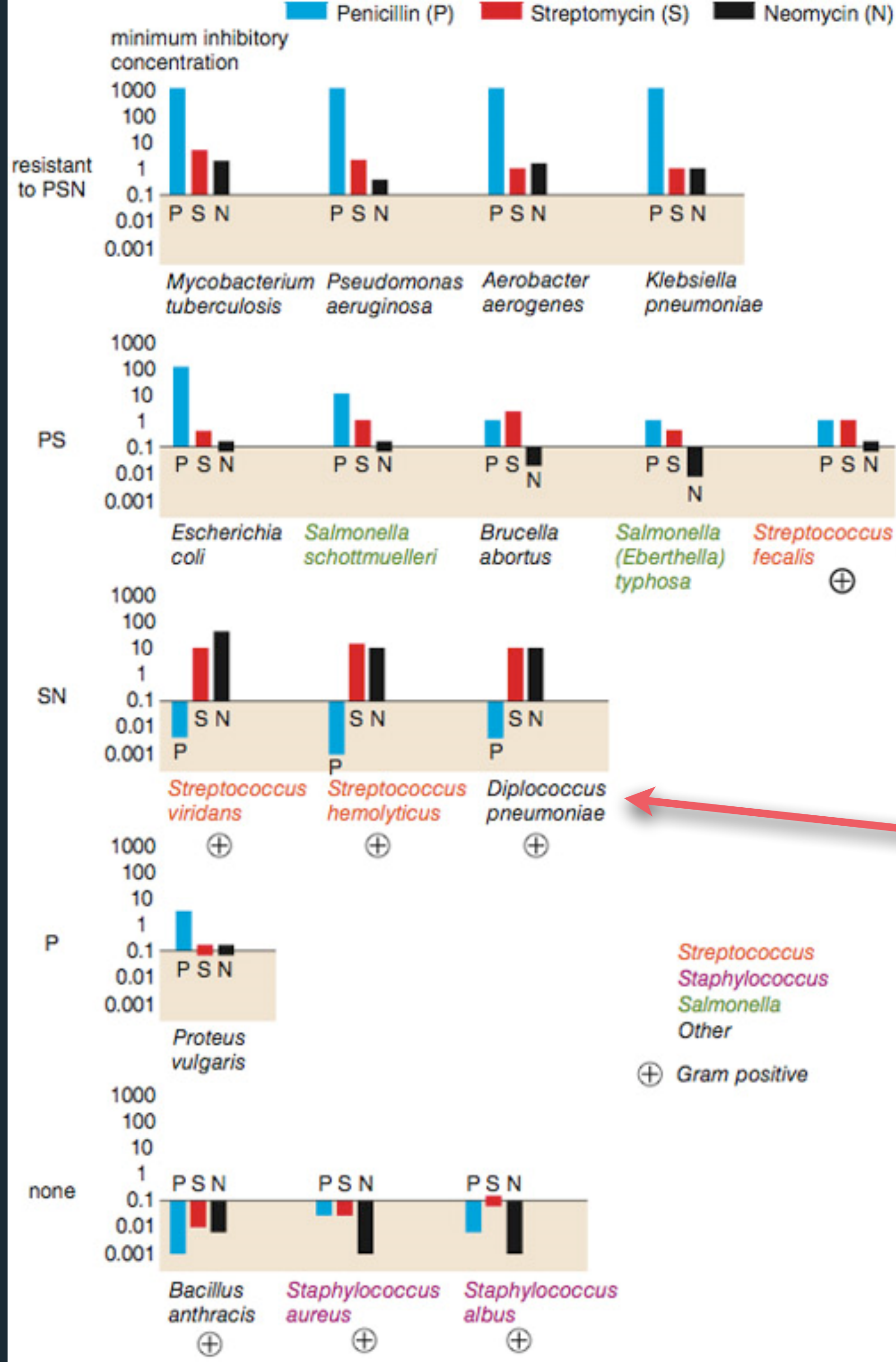
Effectiveness of Antibiotics





Which antibiotic should one use?





Do the bacteria group by antibiotic resistance?

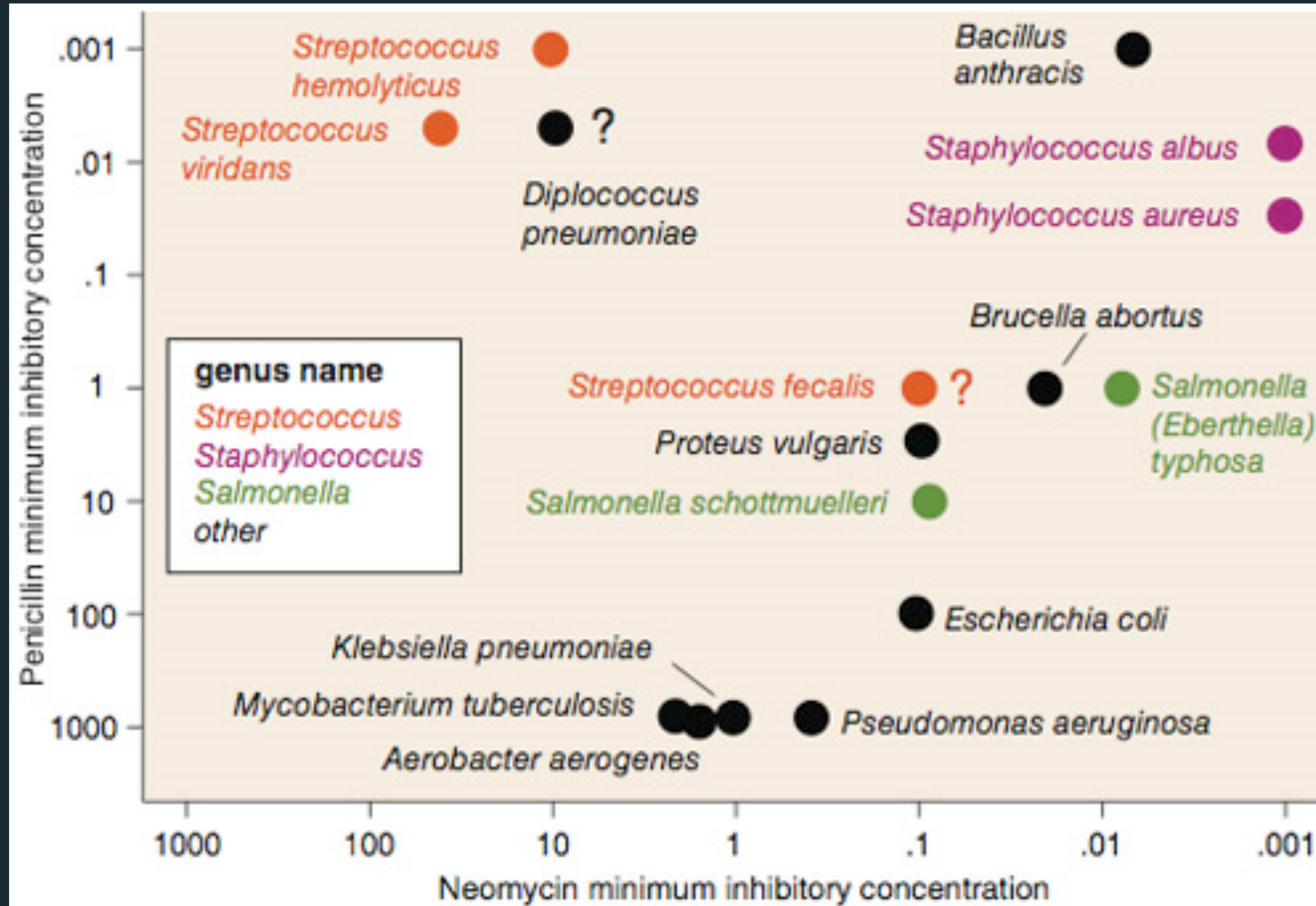
Not a streptococcus!
(realized ~30 yrs later)

Really a streptococcus!
(realized ~20 yrs later)

Do the bacteria group by resistance?

Do different drugs correlate?

Wainer & Lysen. American Scientist, 2009



Exploratory Visual Analysis




Process

1. Construct graphics to address questions.
2. Inspect "answer" and ask new questions.
3. Iterate...

Lessons



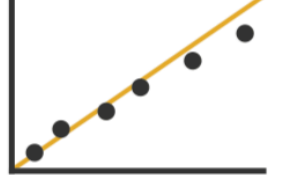

- ✓ Check **data quality** and your **assumptions**.
- ✓ Start with **univariate summaries**, then consider **relationships between variables**.
- ✓ Avoid **premature fixation**: balance **data variation** and **design variation**.

→ Trends → Outliers → Features






→ One → Many





→ Distribution → Dependency → Correlation → Similarity

→ Extremes

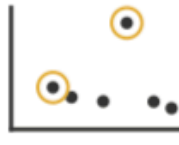




→ Search

	Target known	Target unknown
Location known	 <i>Lookup</i>	 <i>Browse</i>
Location unknown	 <i>Locate</i>	 <i>Explore</i>

→ Query

→ Identify → Compare → Summarize

PART TWO

Exploratory Visual Analysis

PART THREE

Visualizing Uncertainty

The possibility of many/other outcomes.

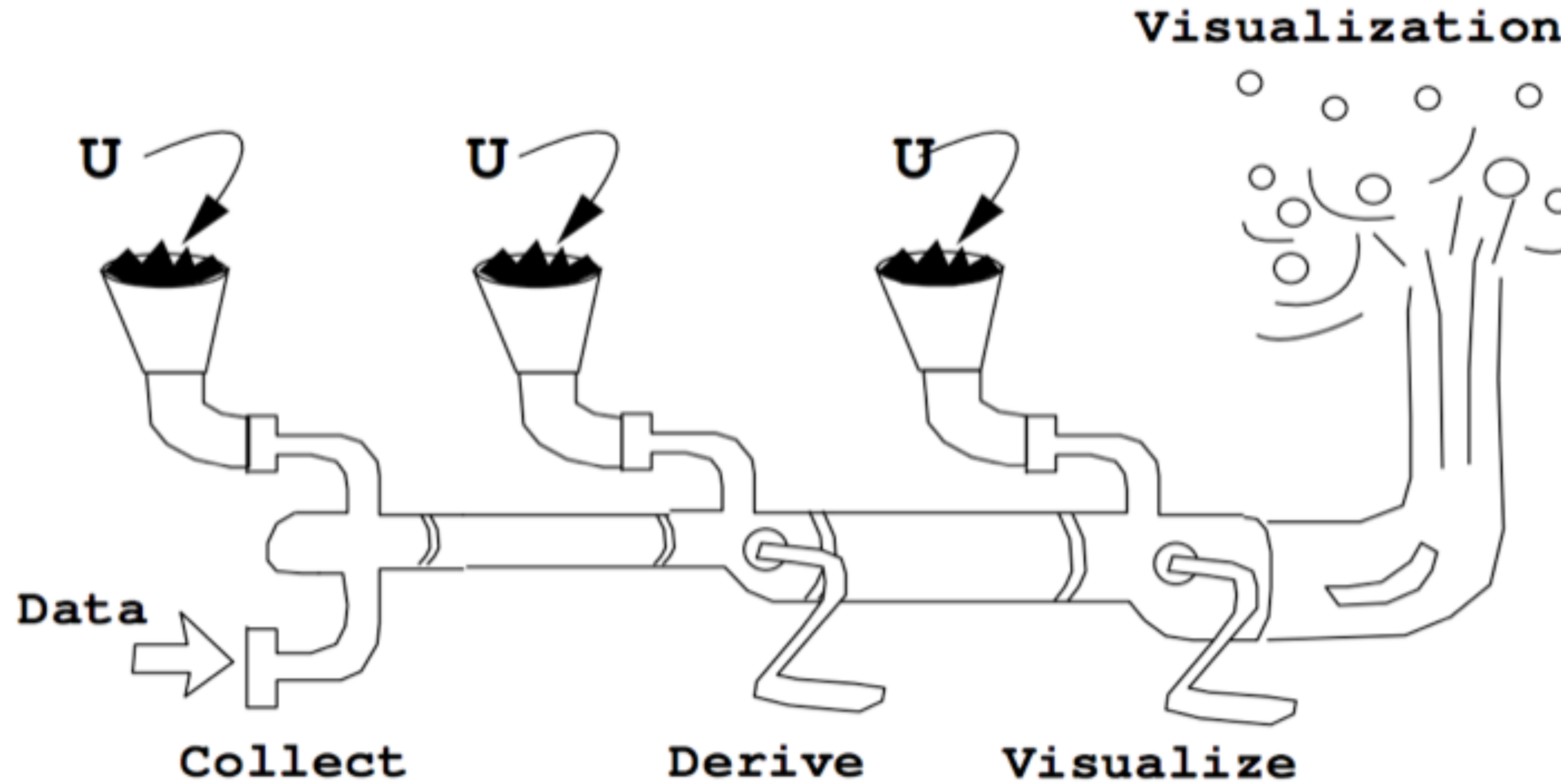
PART THREE

Visualizing Uncertainty



Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

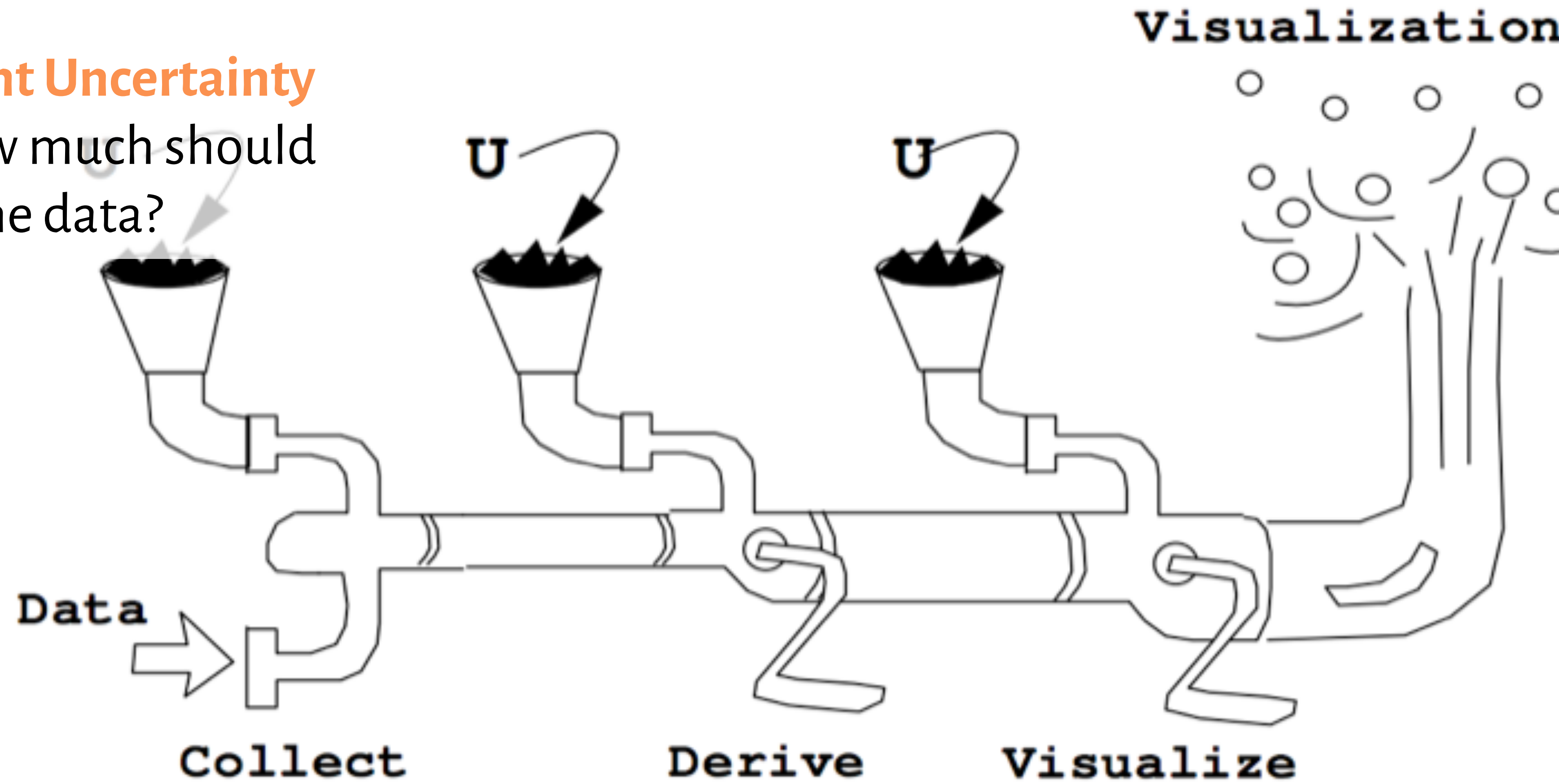


Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

Measurement Uncertainty

How and how much should we sample the data?



Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

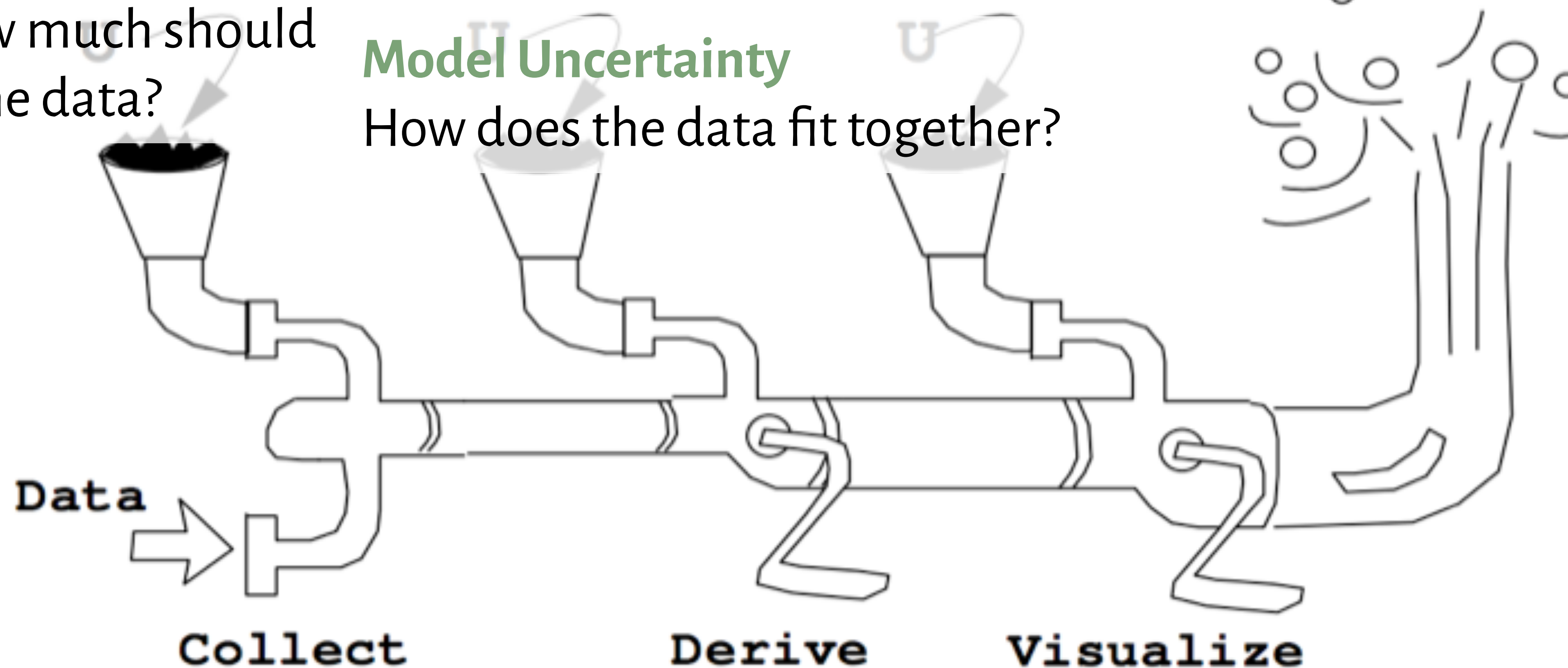
Measurement Uncertainty

How and how much should we sample the data?

Model Uncertainty

How does the data fit together?

Visualization



Sources & Types of *Uncertainty*

The possibility of many/other outcomes.

Measurement Uncertainty

How and how much should we sample the data?

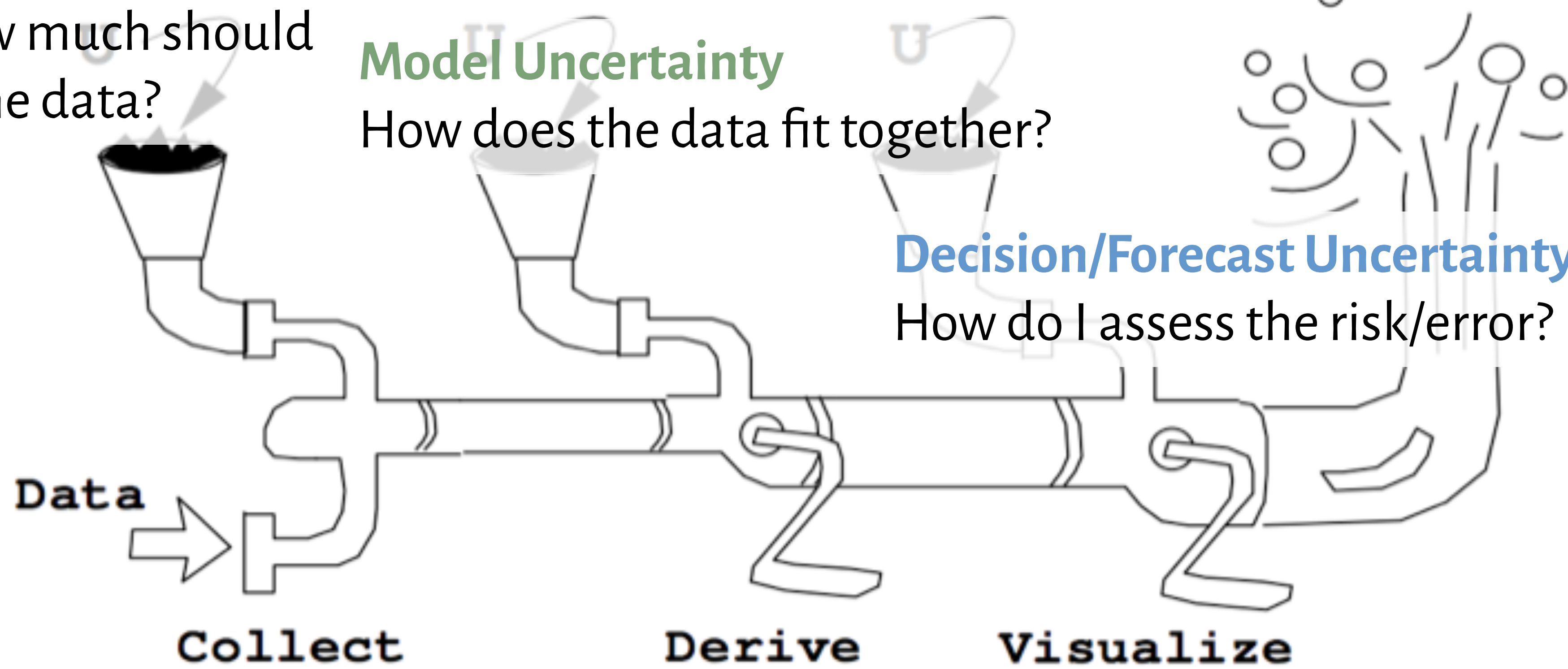
Model Uncertainty

How does the data fit together?

Decision/Forecast Uncertainty

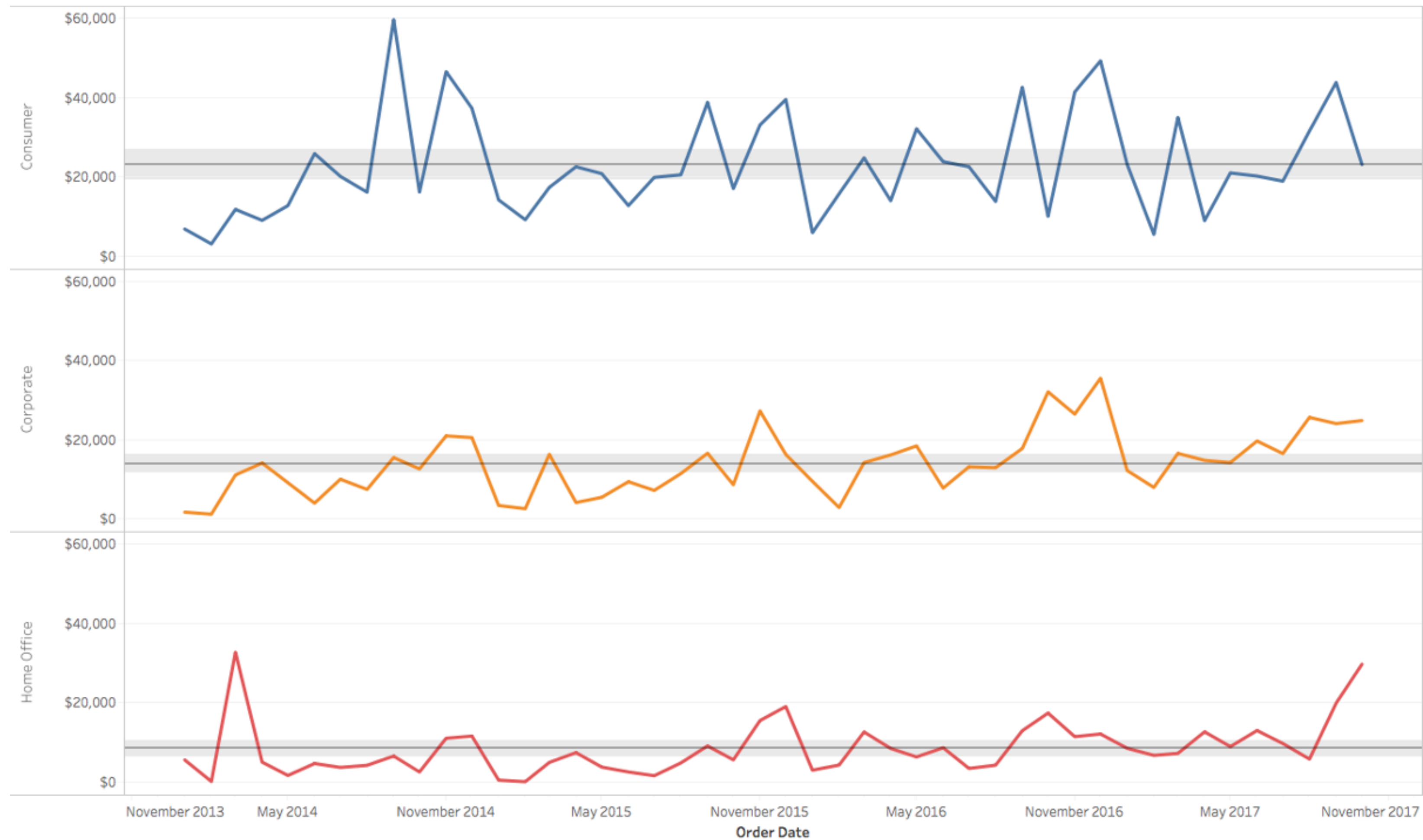
How do I assess the risk/error?

Visualization



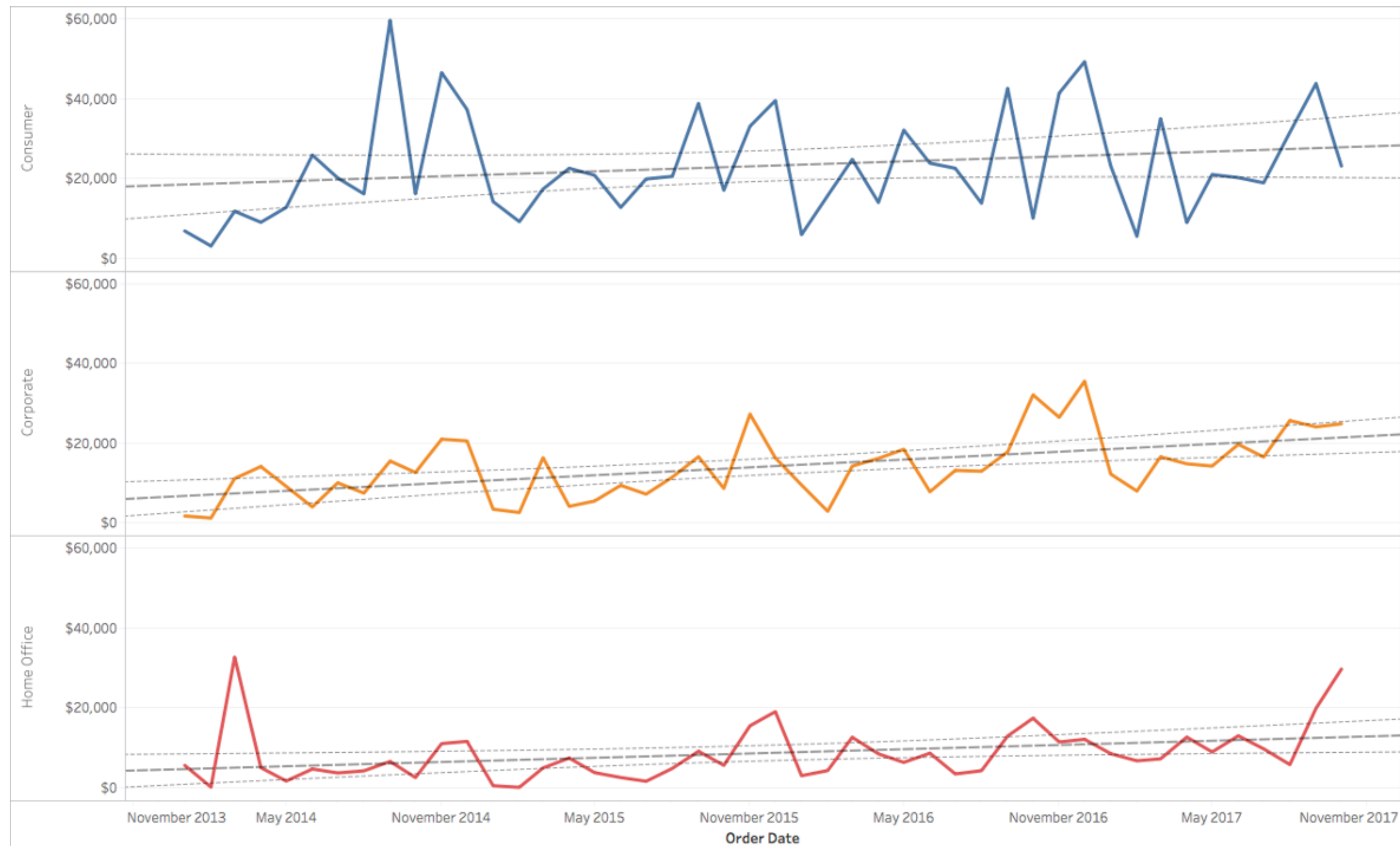
Visualizing Uncertainty

Measurement uncertainty in Tableau



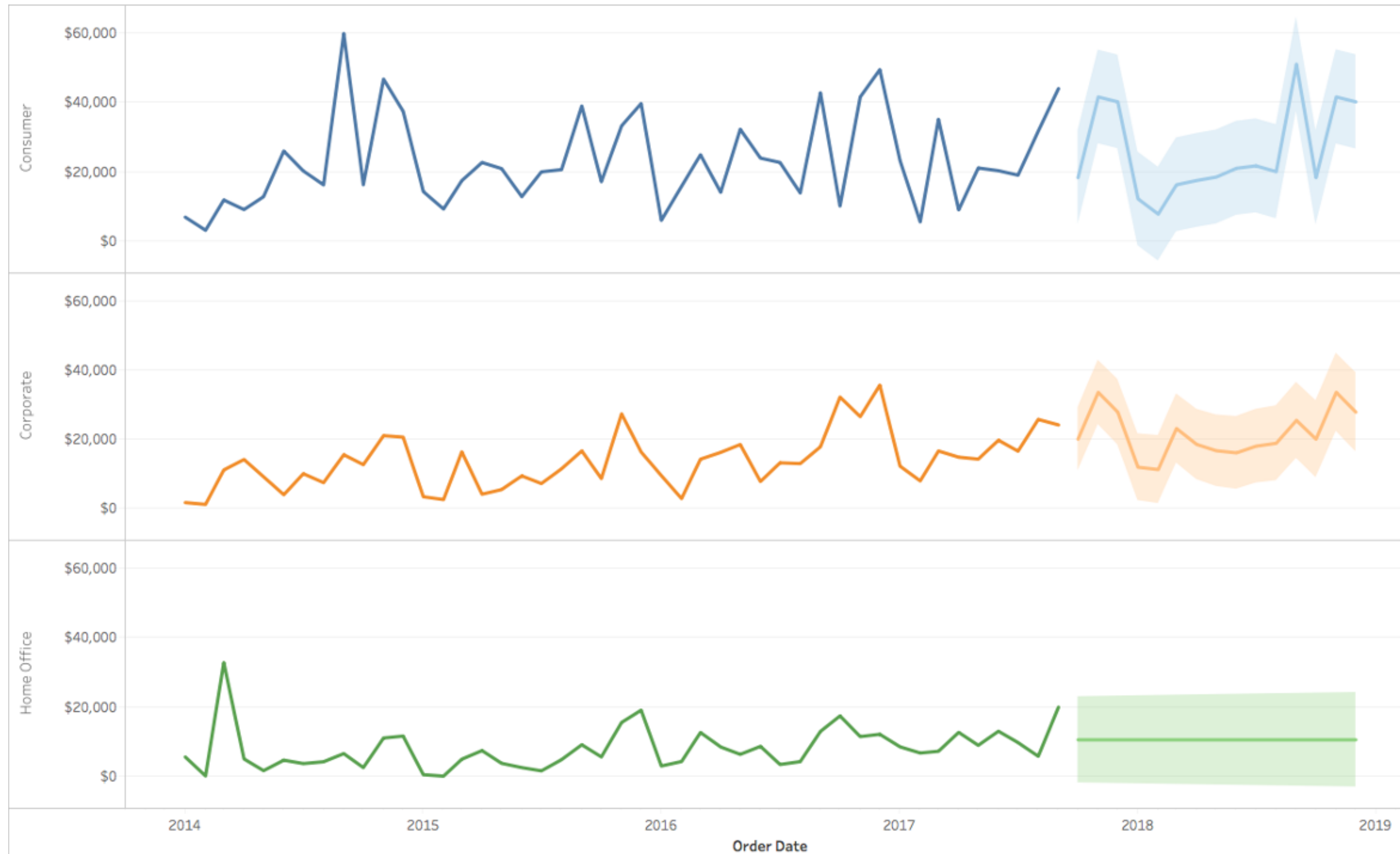
Visualizing Uncertainty

Model uncertainty in Tableau

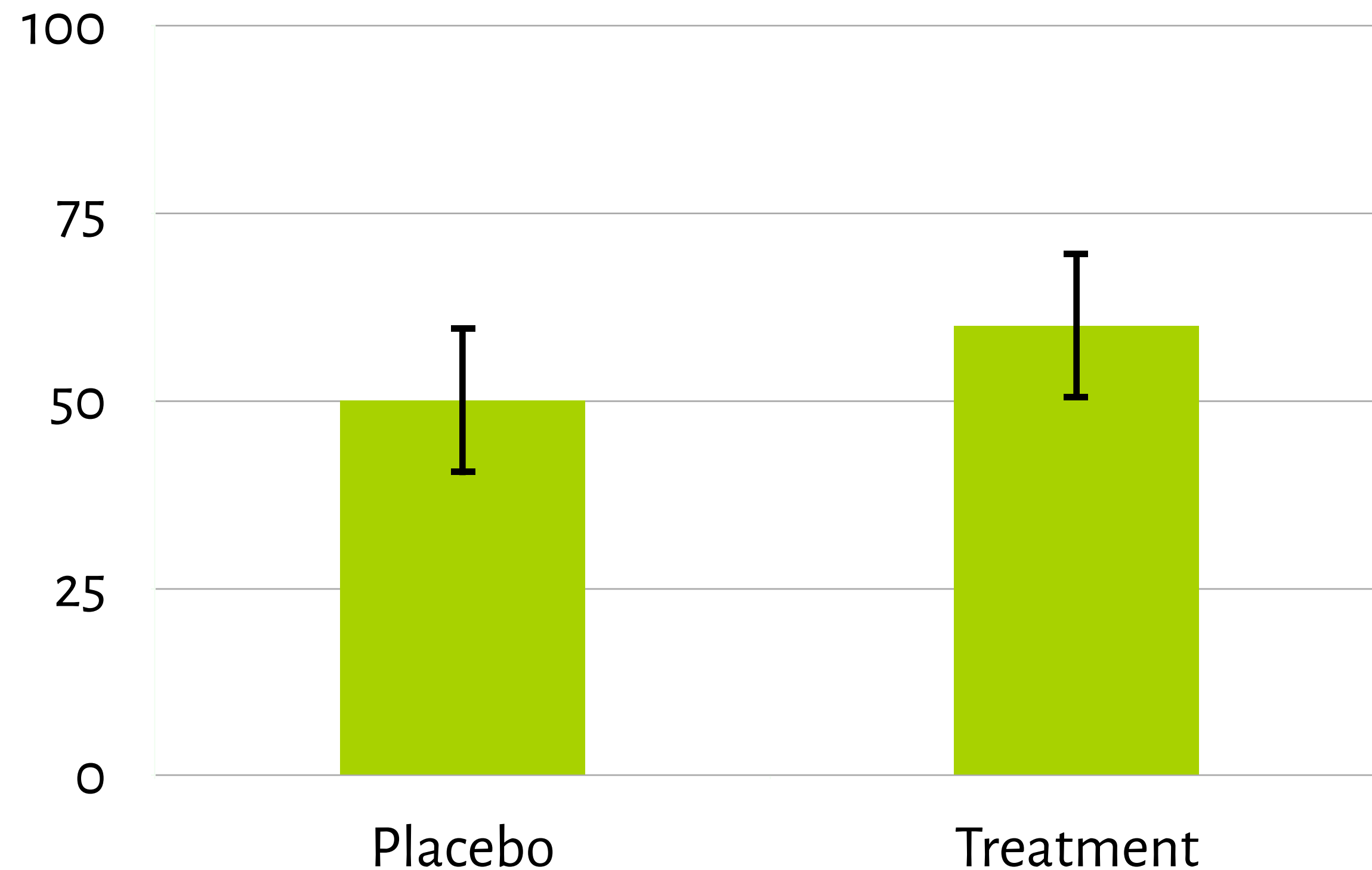


Visualizing Uncertainty

Forecast uncertainty in Tableau

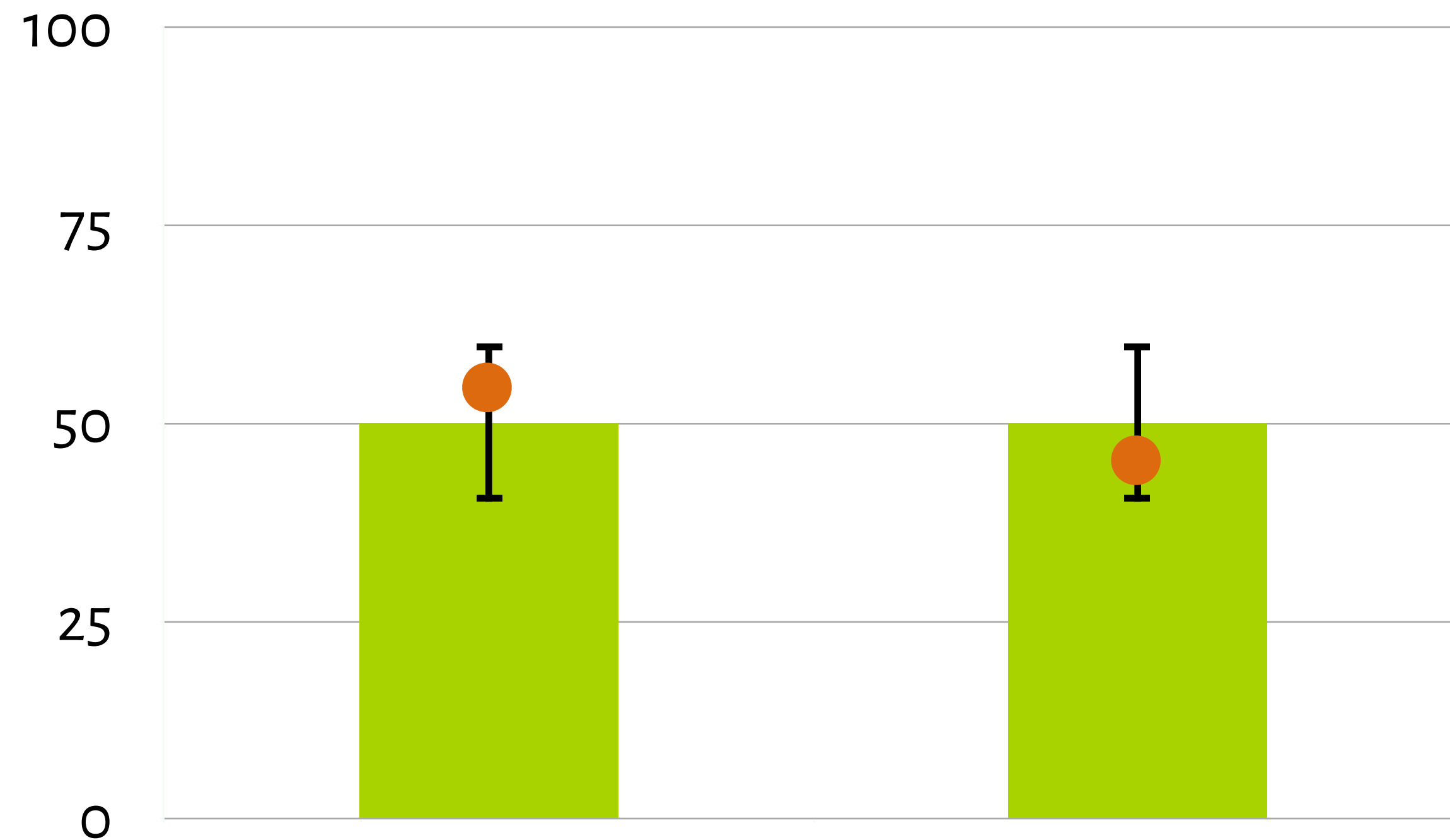


Visualizing Uncertainty: Glyphs



✘ Error bars **aren't consistently used** to visualize the same measure (standard deviation, standard error, IQR, 95% CI).

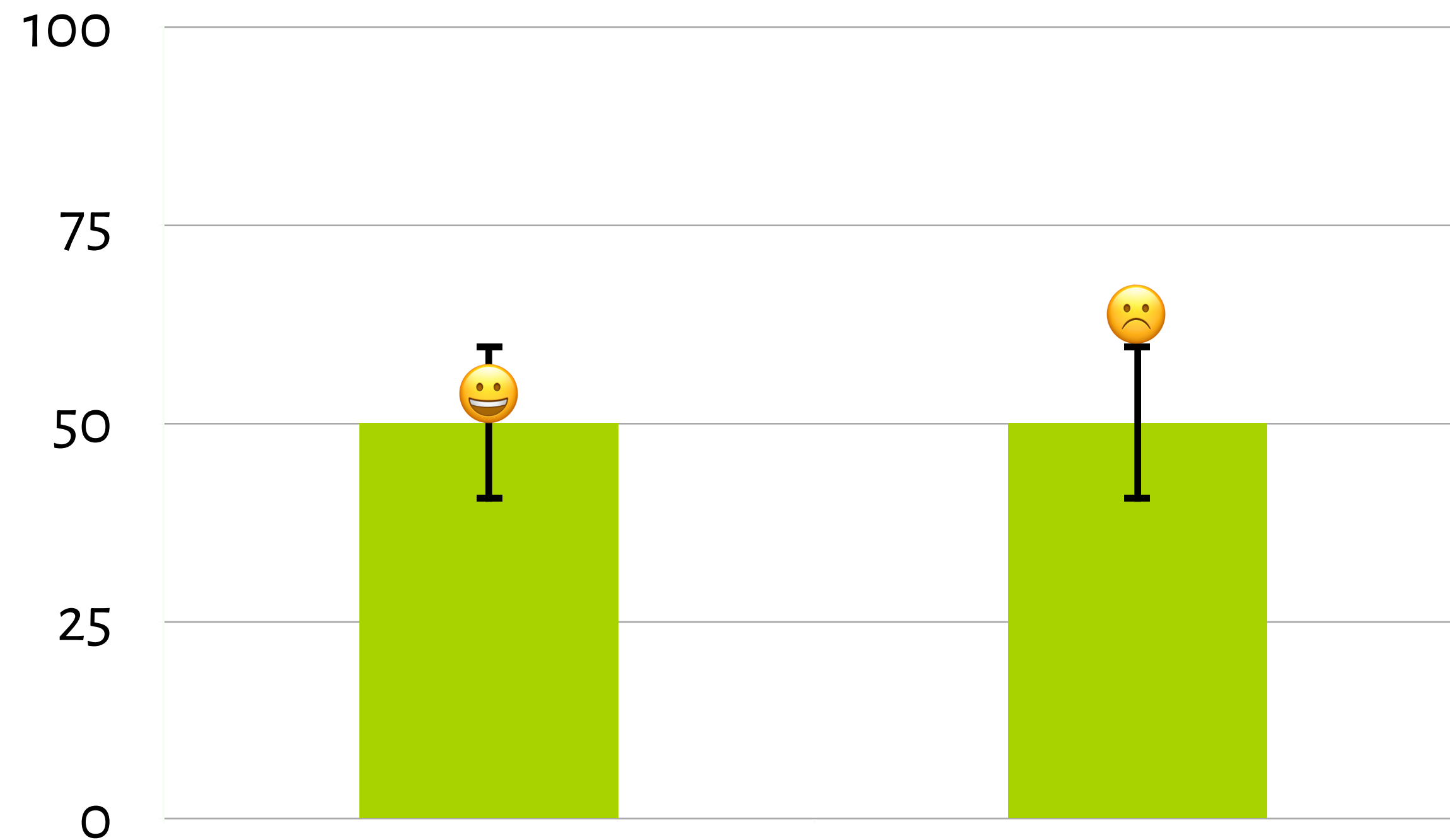
Visualizing Uncertainty: Glyphs



- ✘ Error bars **aren't consistently used** to visualize the same measure (standard deviation, standard error, IQR, 95% CI).
- ✘ **Within-the-bar bias**: people perceive points falling within the bar as more likely than those that lie outside.

[Newman & Scholl, 2012]
[Correll & Gleicher, 2014]

Visualizing Uncertainty: Glyphs

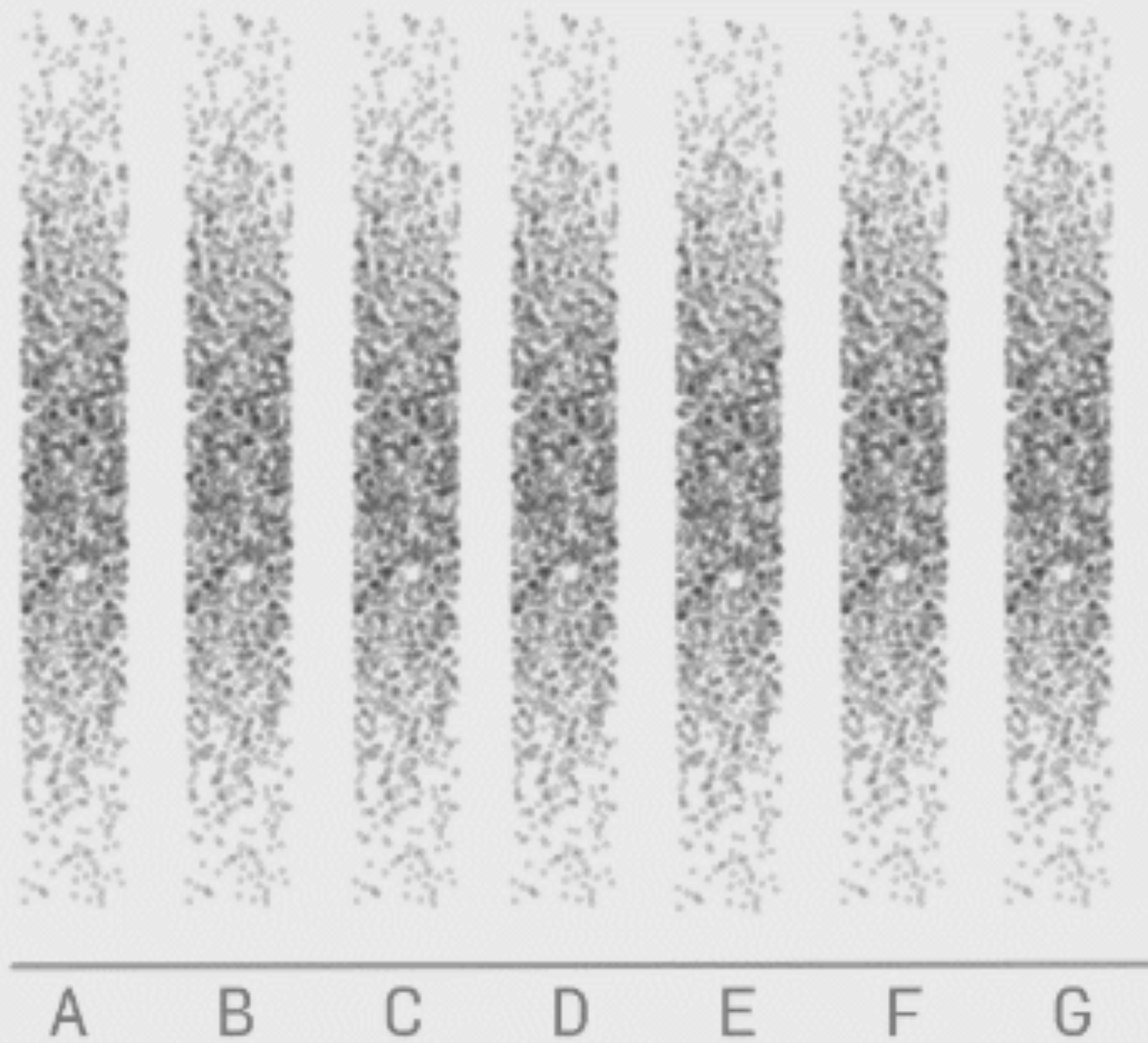


- ✘ Error bars **aren't consistently used** to visualize the same measure (standard deviation, standard error, IQR, 95% CI).
- ✘ **Within-the-bar bias**: people perceive points falling within the bar as more likely than those that lie outside.
- ✘ **Binary bias**: people perceive values to either be in or out of the margins of error.

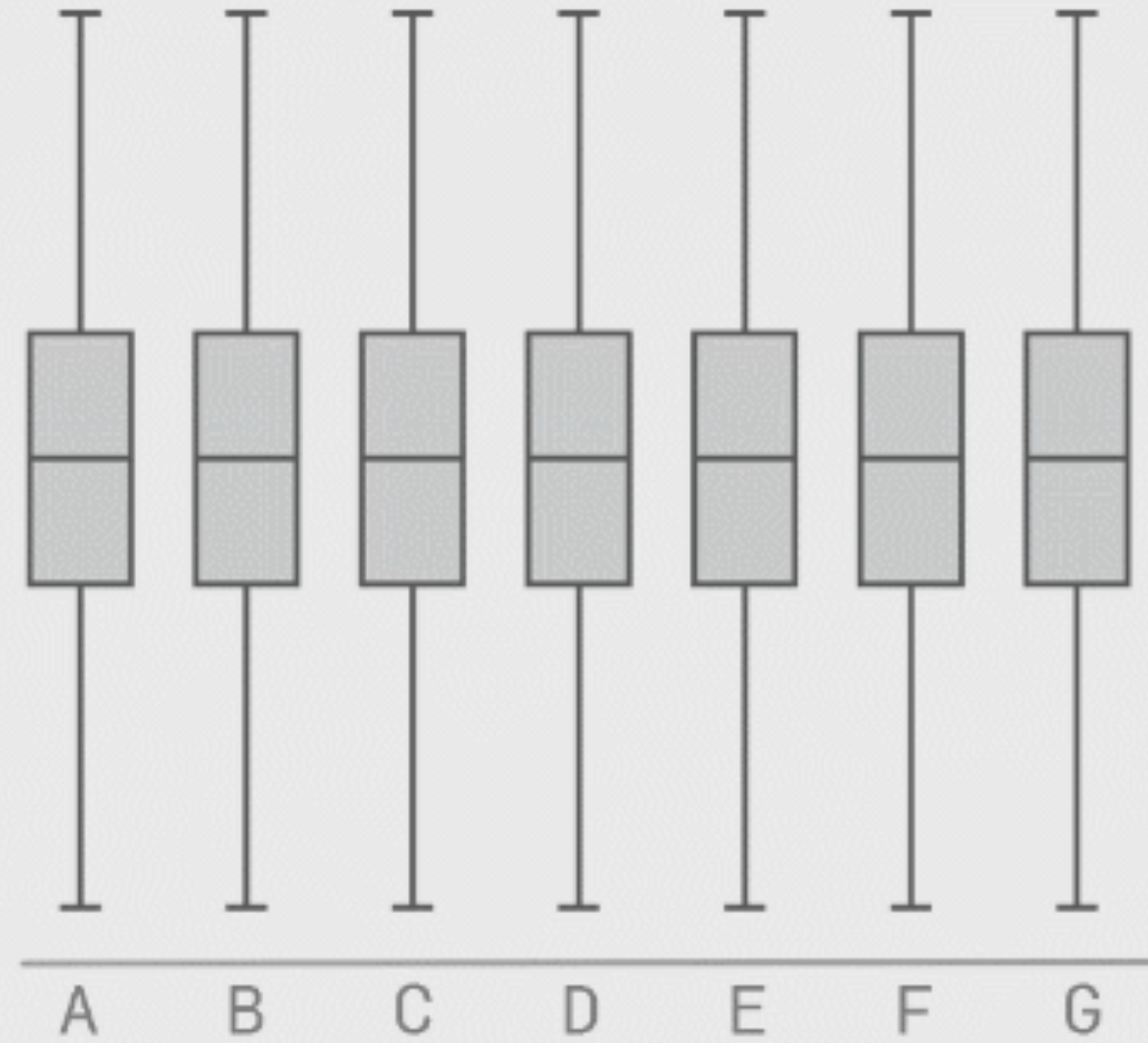
[Newman & Scholl, 2012]
[Correll & Gleicher, 2014]

Visualizing Uncertainty: **Visual Variables**

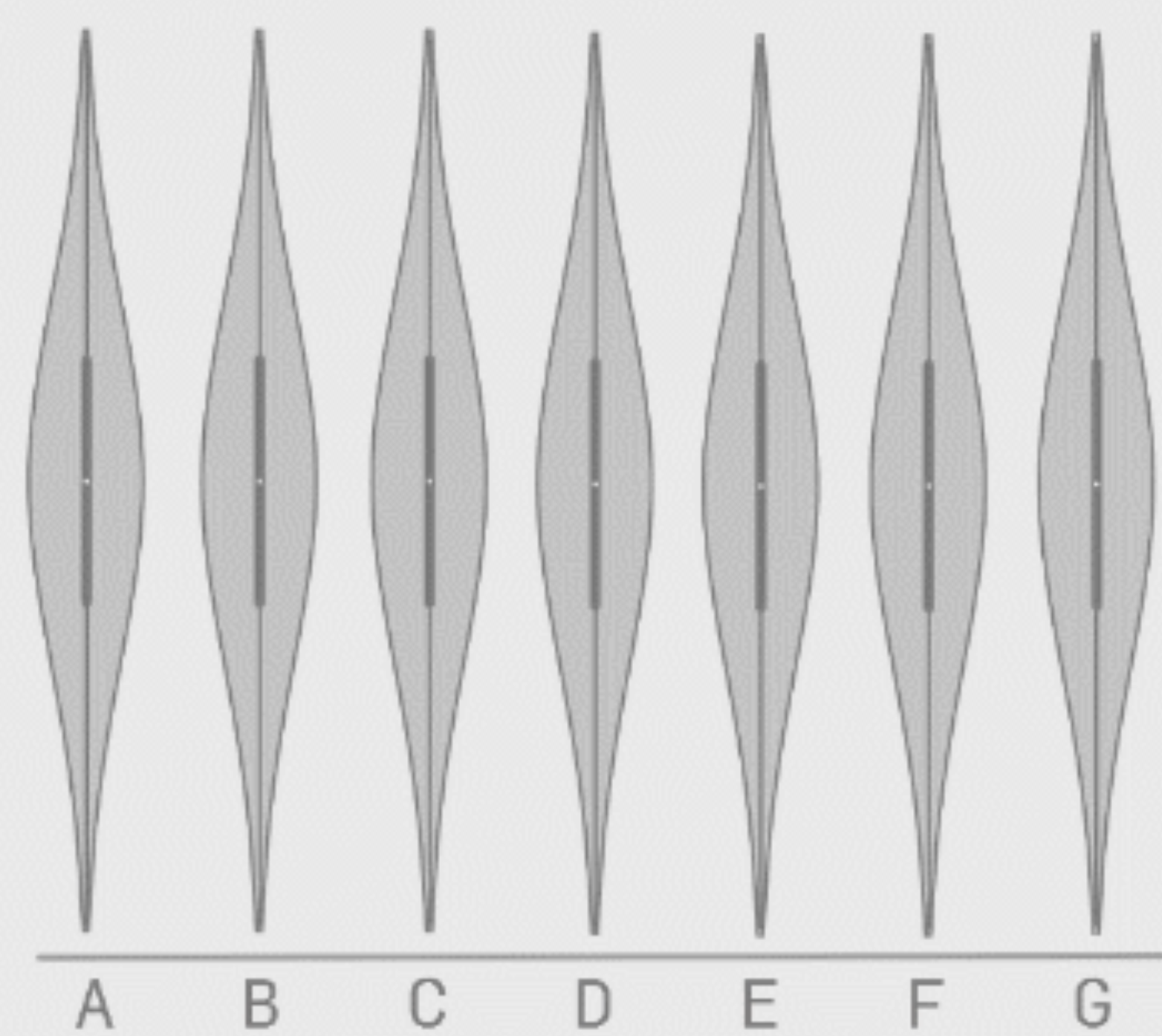
Raw Data



Box-plot of the Data

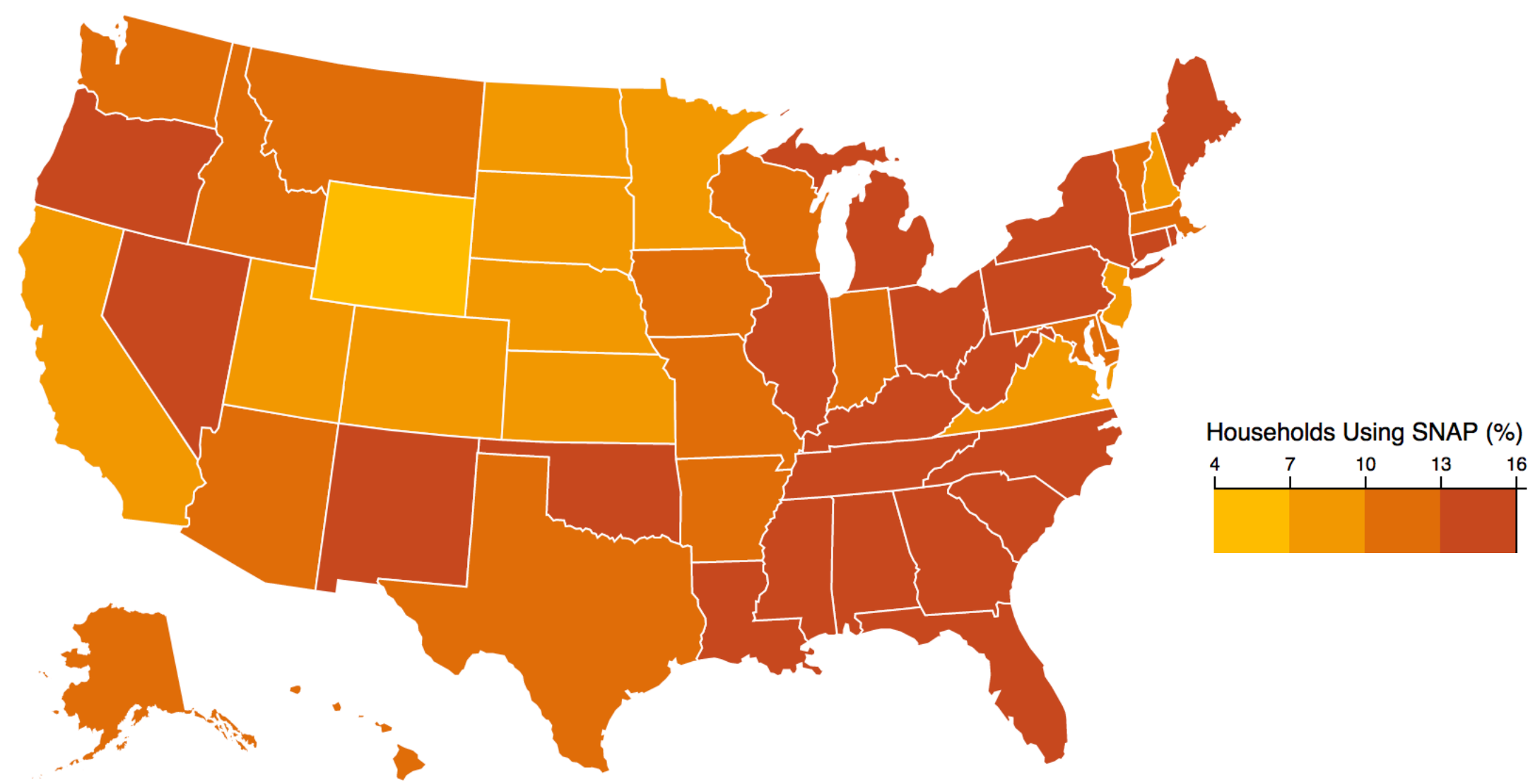


Violin-plot of the Data

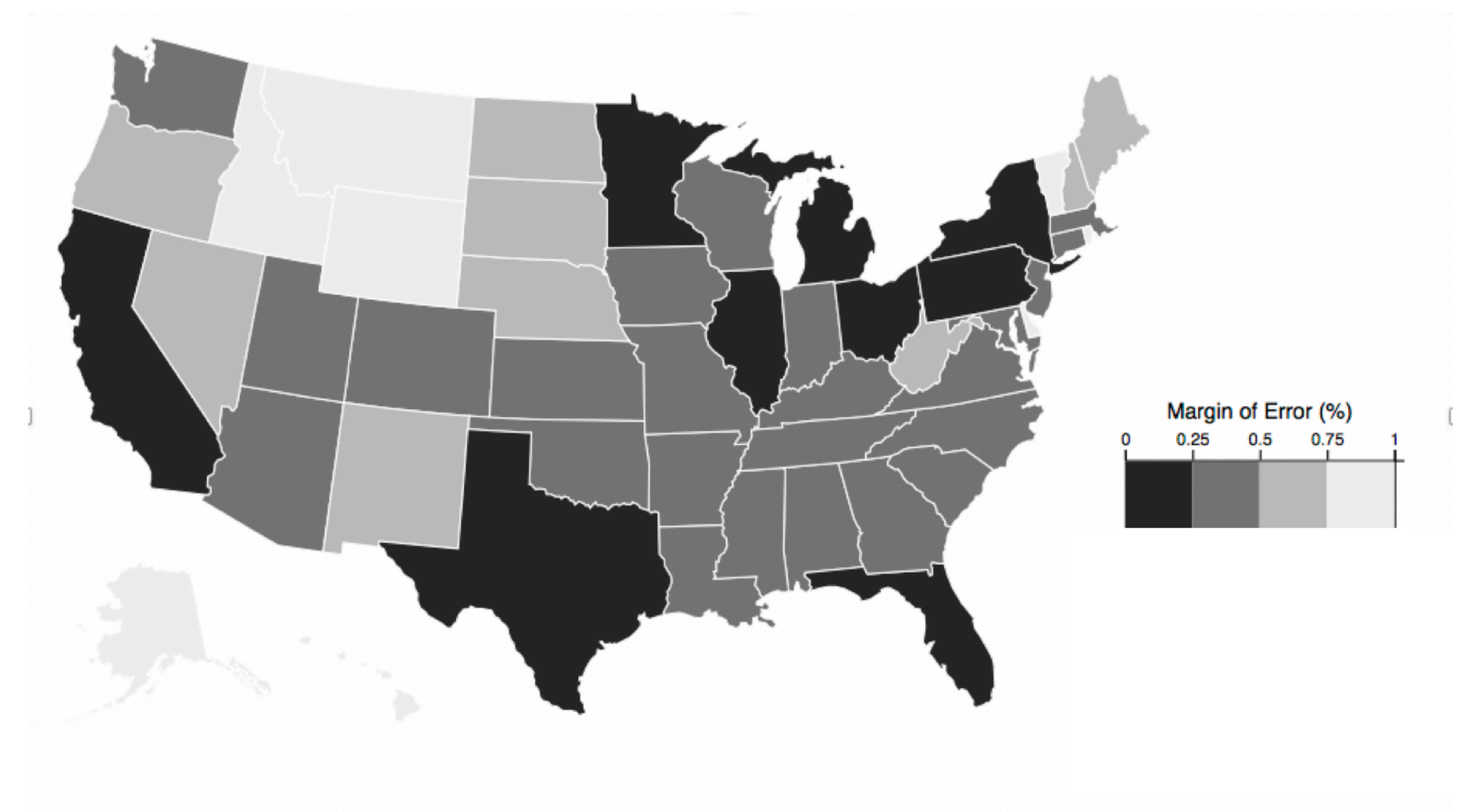


[Matejka & Fitzmaurice, 2017]
[Correll & Gleicher, 2014]

Visualizing Uncertainty: *Visual Variables*



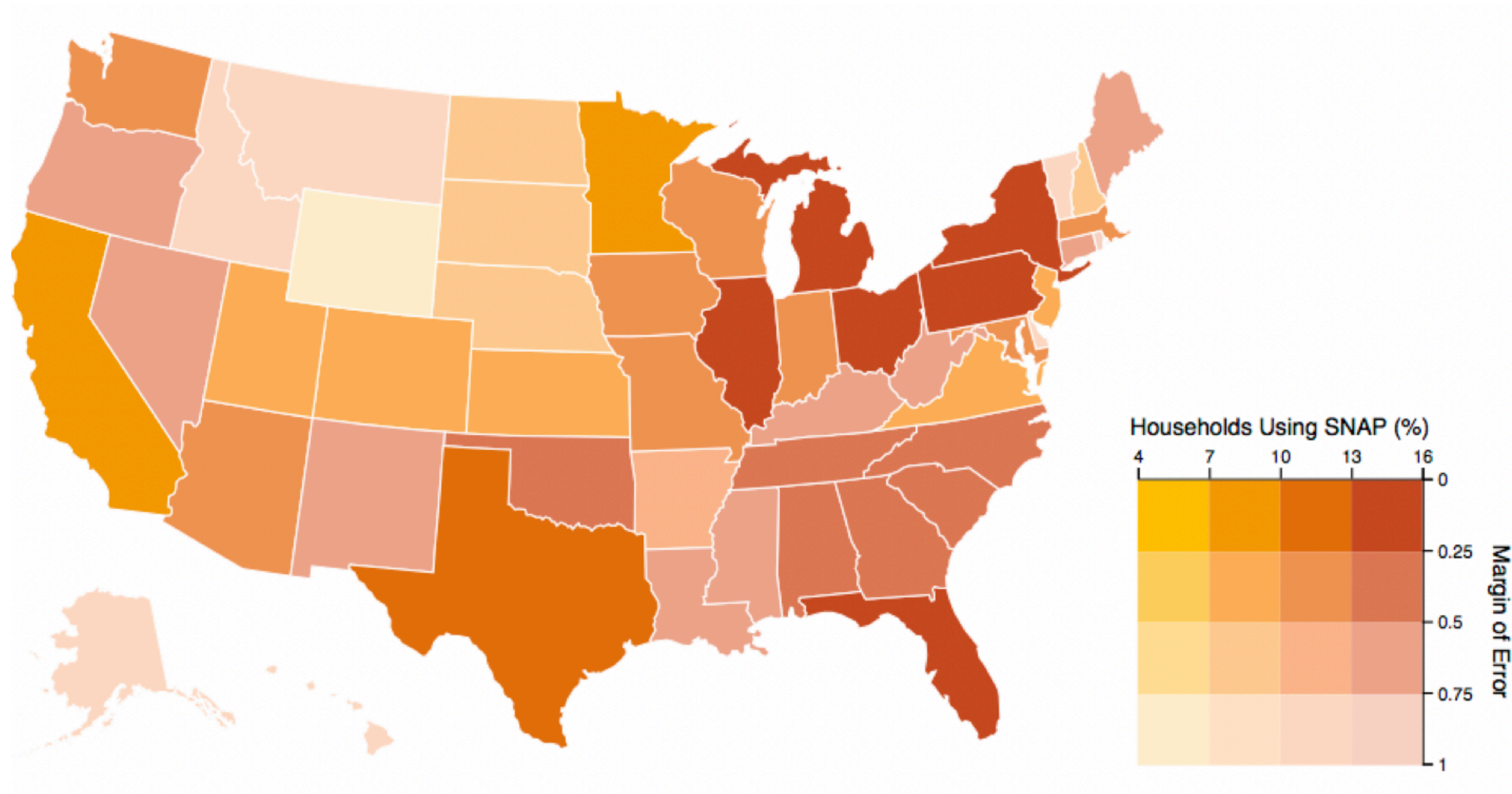
Data Map



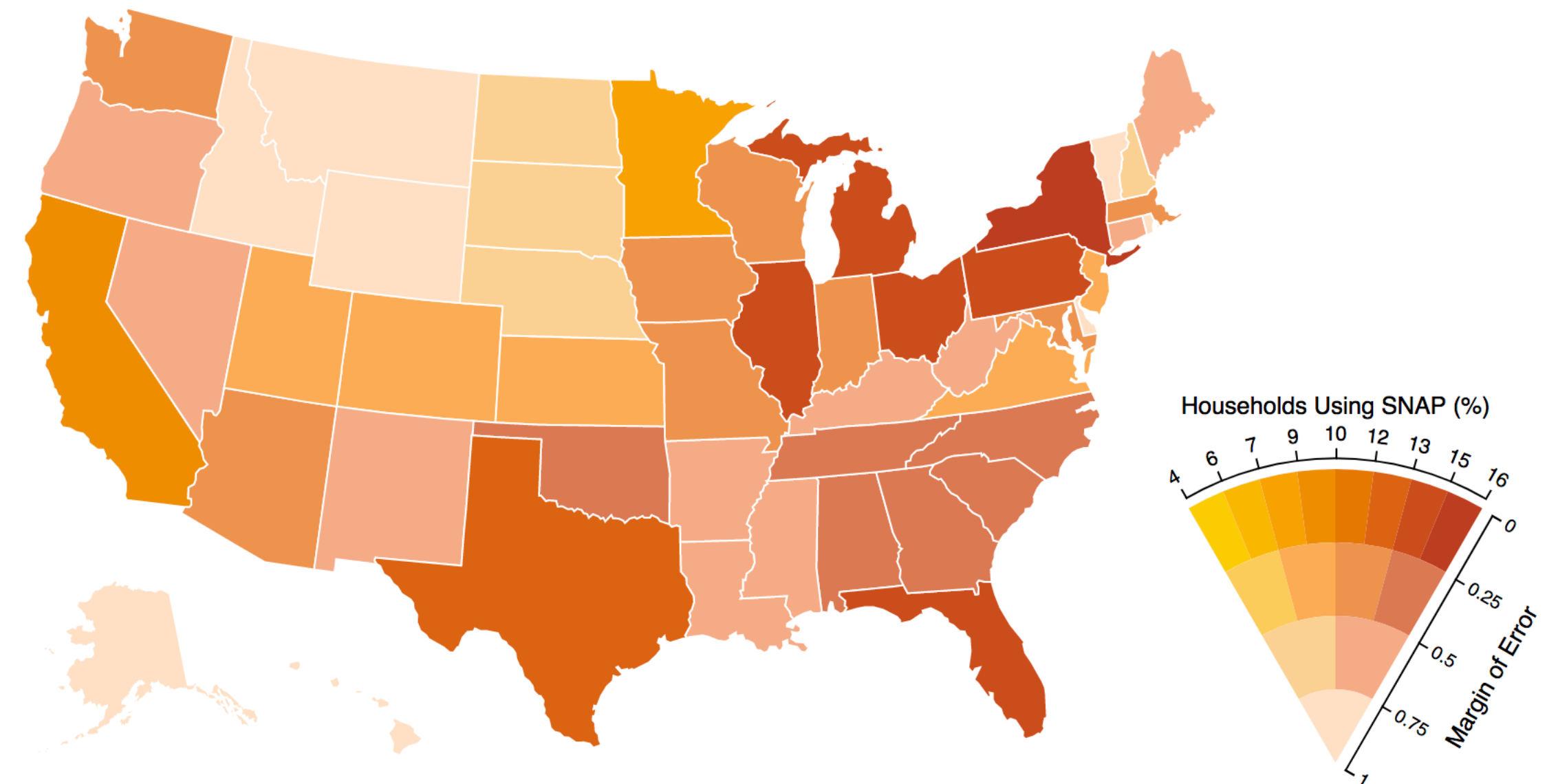
Uncertainty Map

[Correll, Moritz, & Heer, 2018]

Visualizing Uncertainty: *Visual Variables*



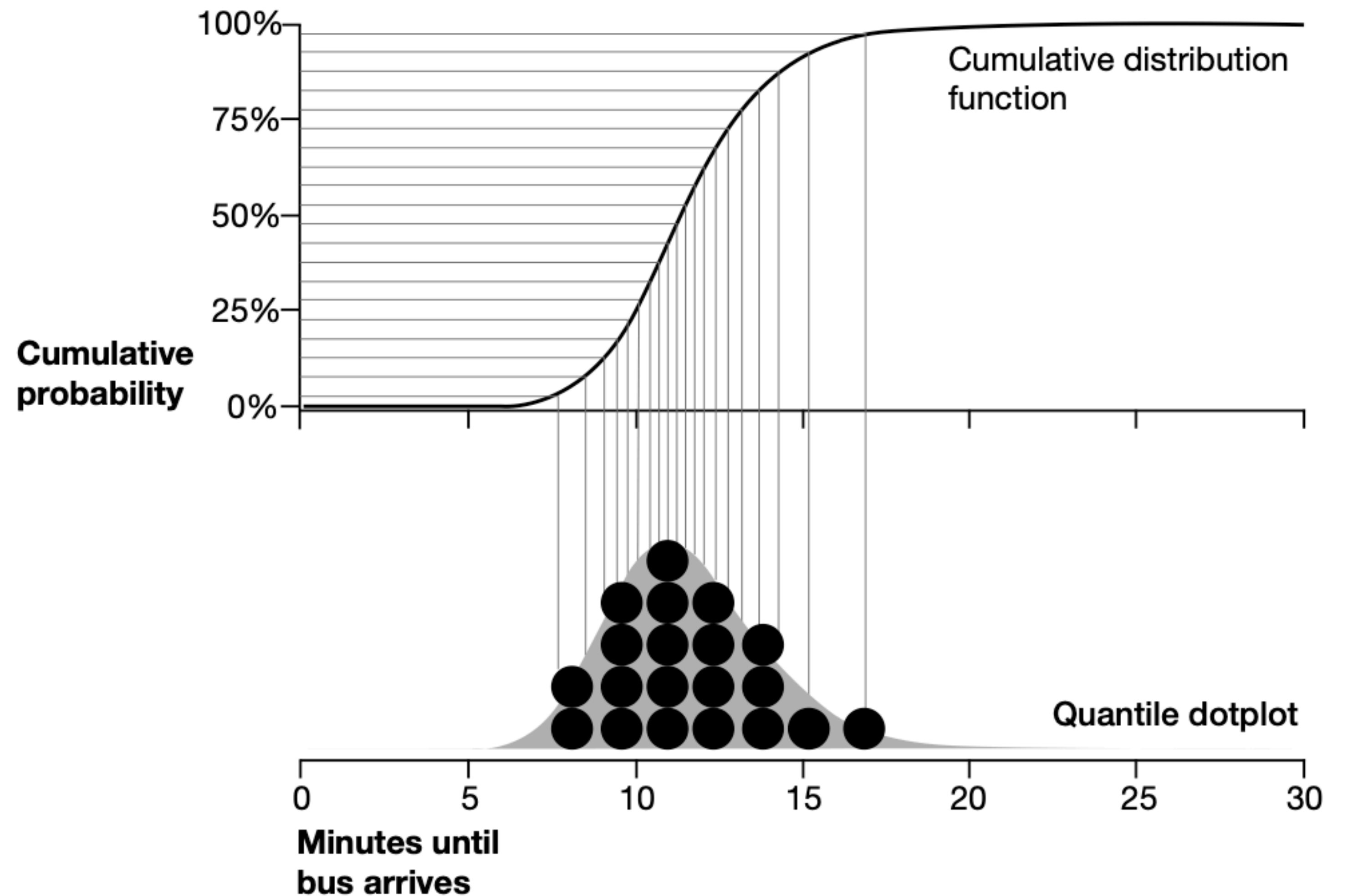
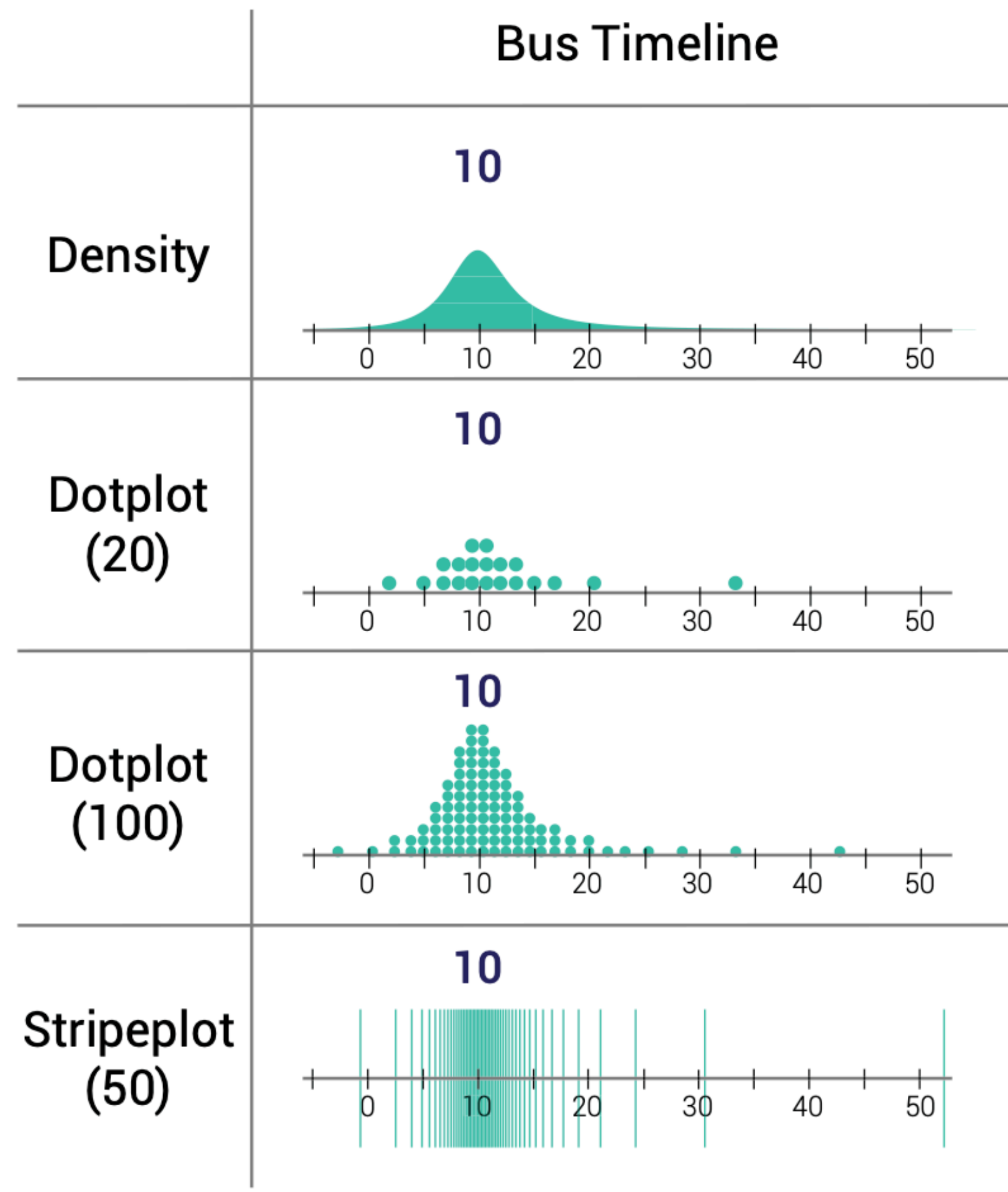
Bivariate Map (Data + Uncertainty)



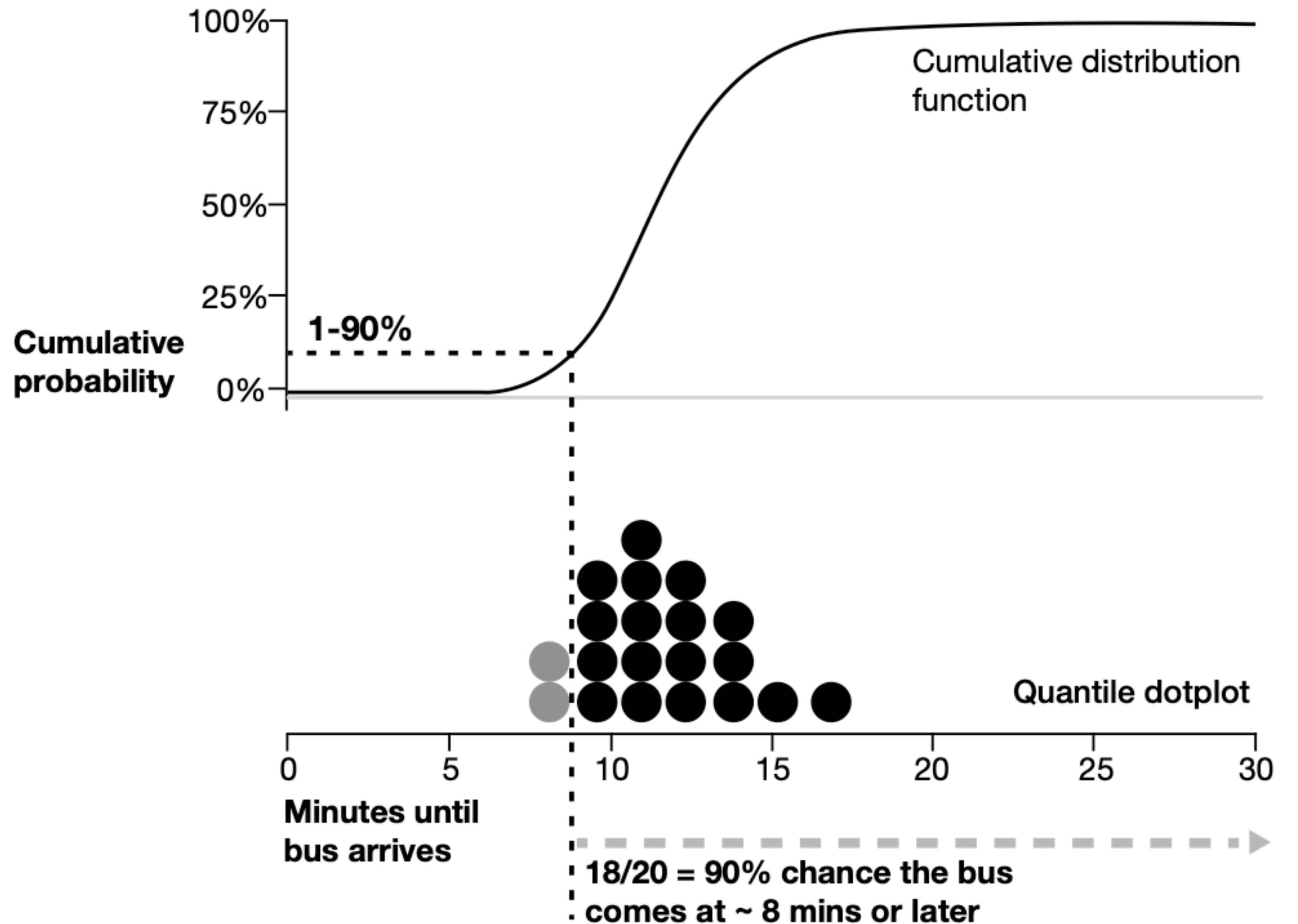
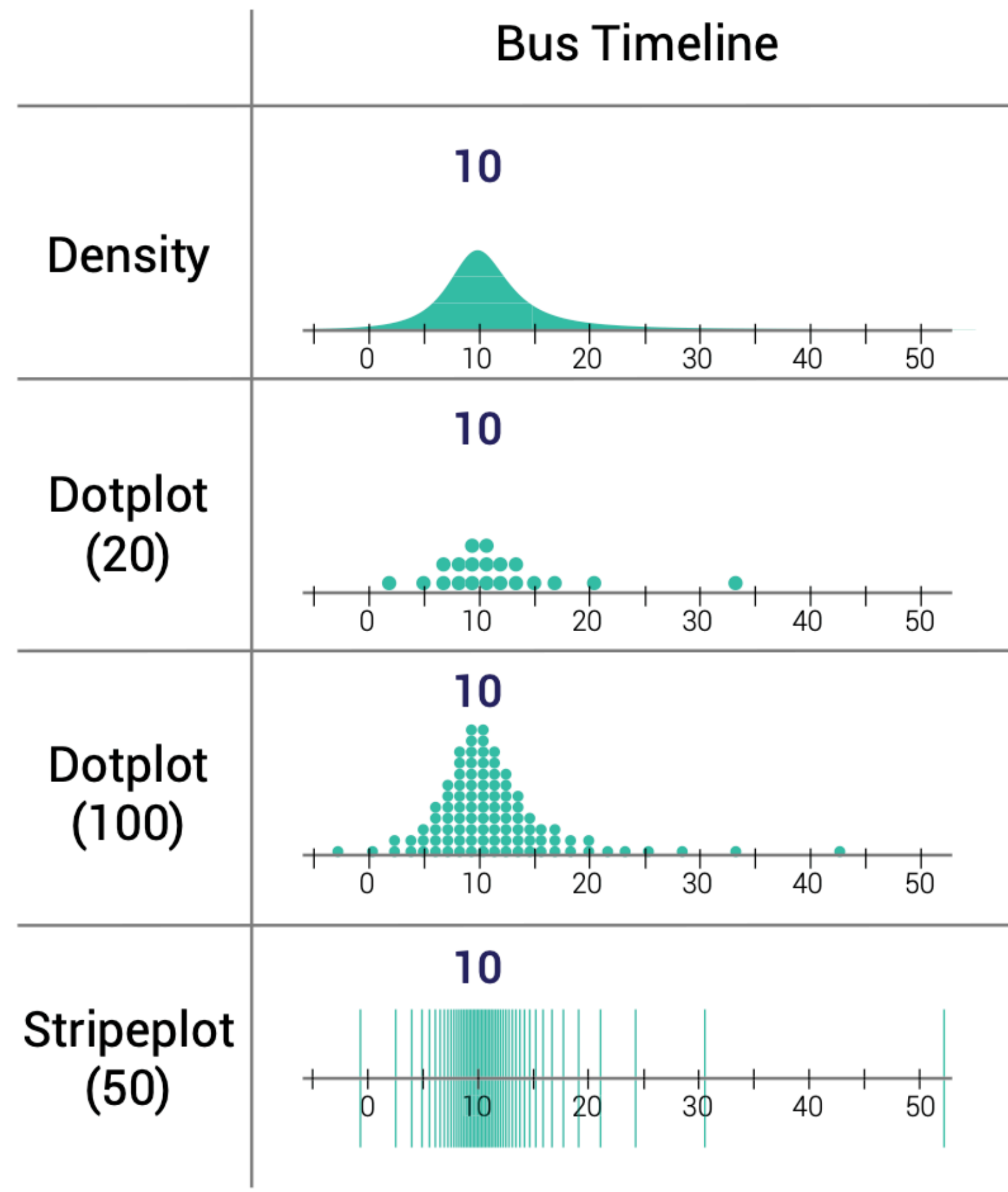
Value-Suppressing Uncertainty Map

[Correll, Moritz, & Heer, 2018]

Visualizing Uncertainty: Set of “Draws”



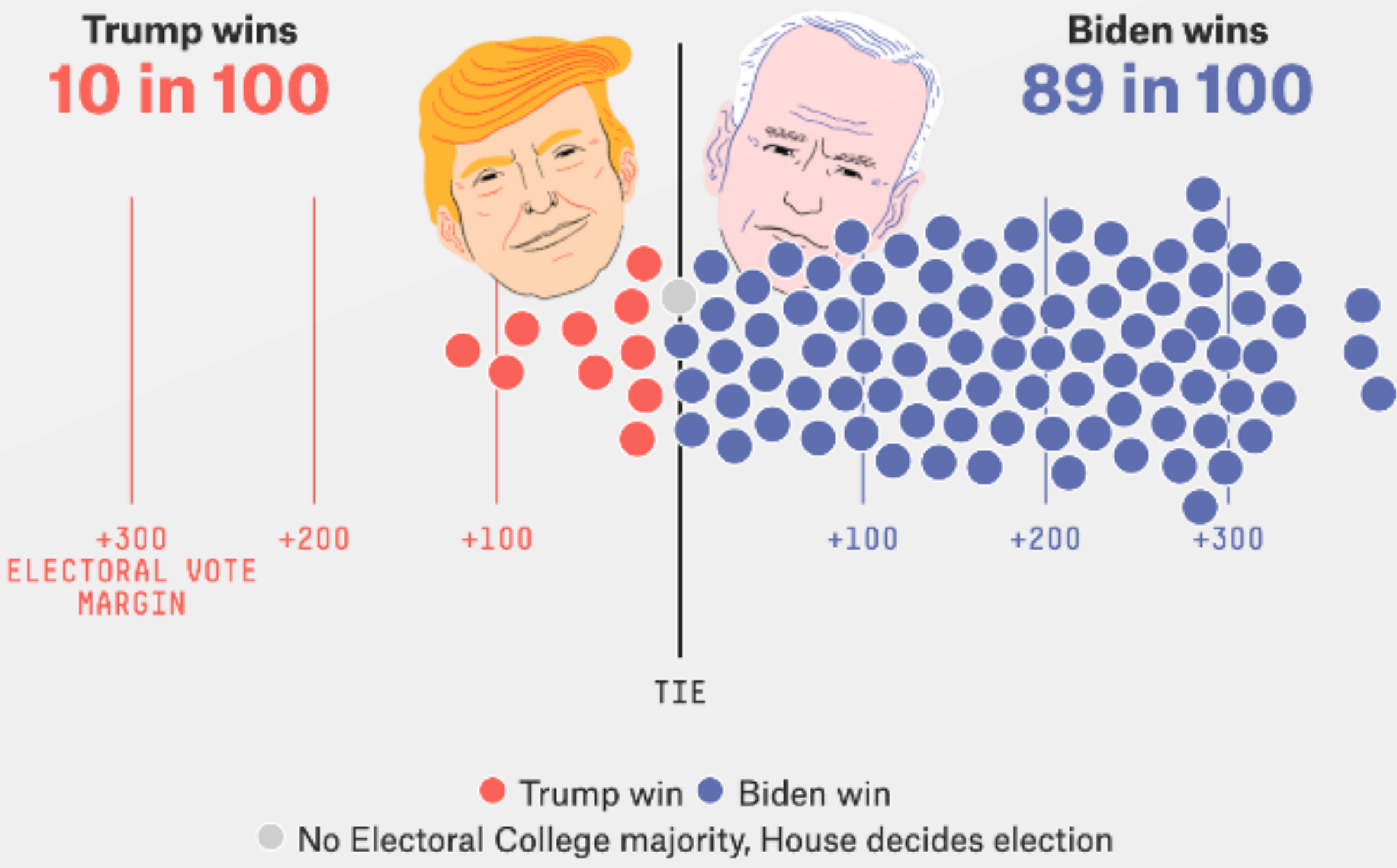
Visualizing Uncertainty: Set of “Draws”



[Kay et al., 2016]

Biden is favored to win the election

We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.



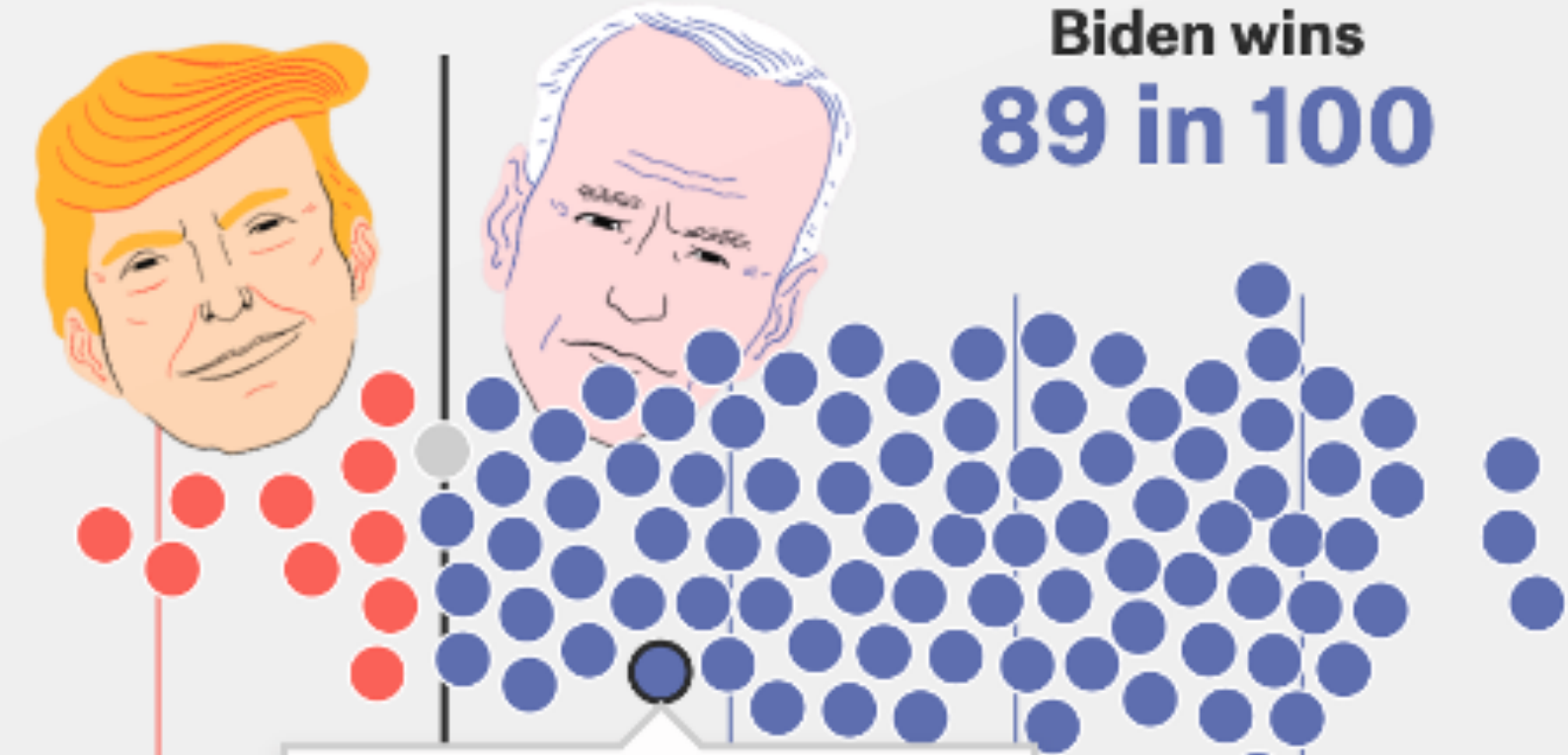
Don't count the underdog out! Upset wins are surprising but not impossible.

Biden is *favored* to win the election

We simulate the election 40,000 times to see who wins most often. The sample of 100 outcomes below gives you a good idea of the range of scenarios our model thinks is possible.

Trump wins
10 in 100

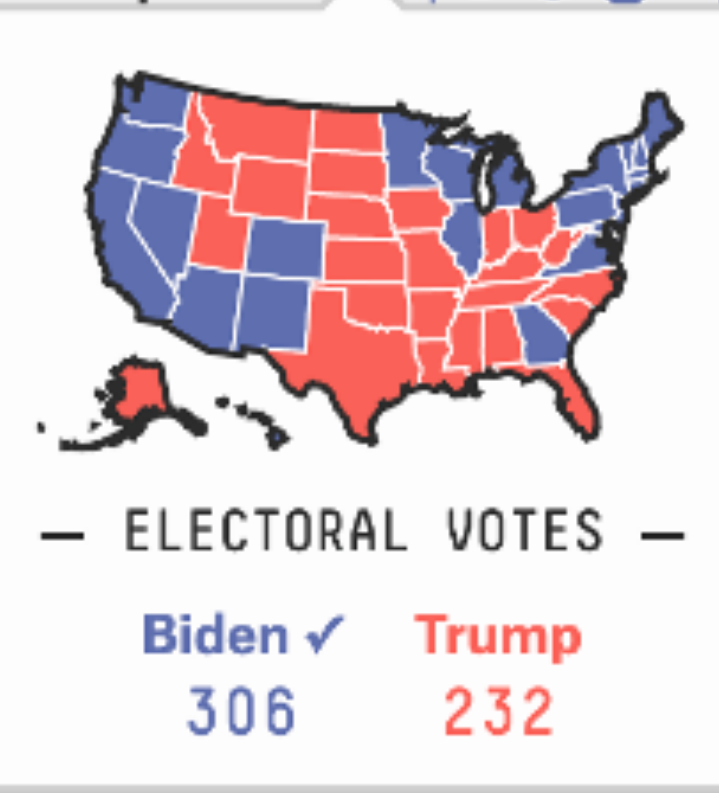
Biden wins
89 in 100



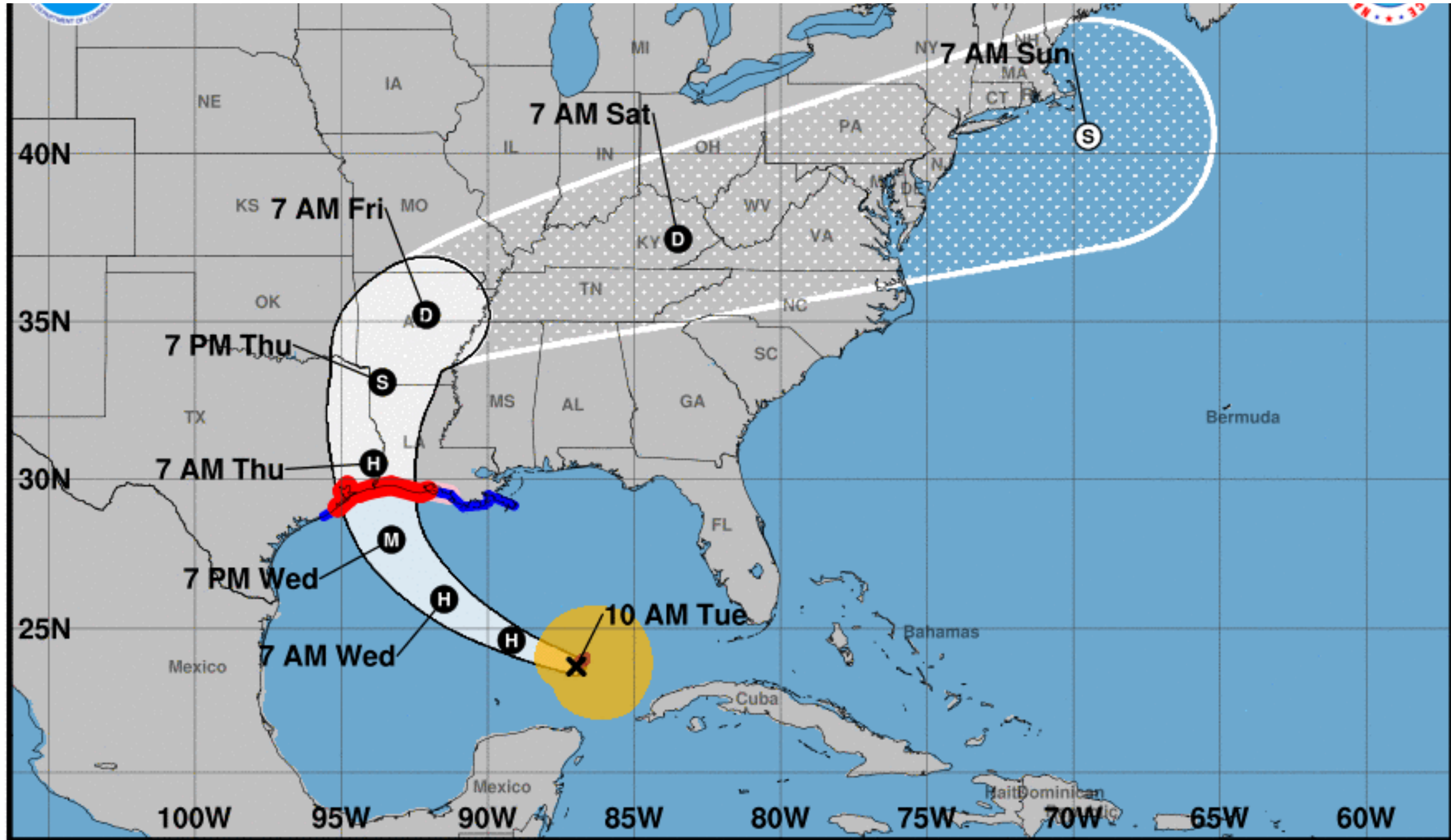
+300
+200
+100
ELECTORAL VOTE MARGIN

0
+300

Don't count the underdog out! Upset wins are surprising but not impossible.



● Trump
● No Electoral College



Hurricane Laura
 Tuesday August 25, 2020
 10 AM CDT Advisory 23
 NWS National Hurricane Center

Current information: x
 Center location 23.7 N 87.0 W
 Maximum sustained wind 75 mph
 Movement WNW at 16 mph

Forecast positions:
 ● Tropical Cyclone ○ Post/Potential TC
 Sustained winds: D < 39 mph
 S 39-73 mph H 74-110 mph M > 110 mph

Potential track area: Day 1-3 (solid line) Day 4-5 (dotted line)

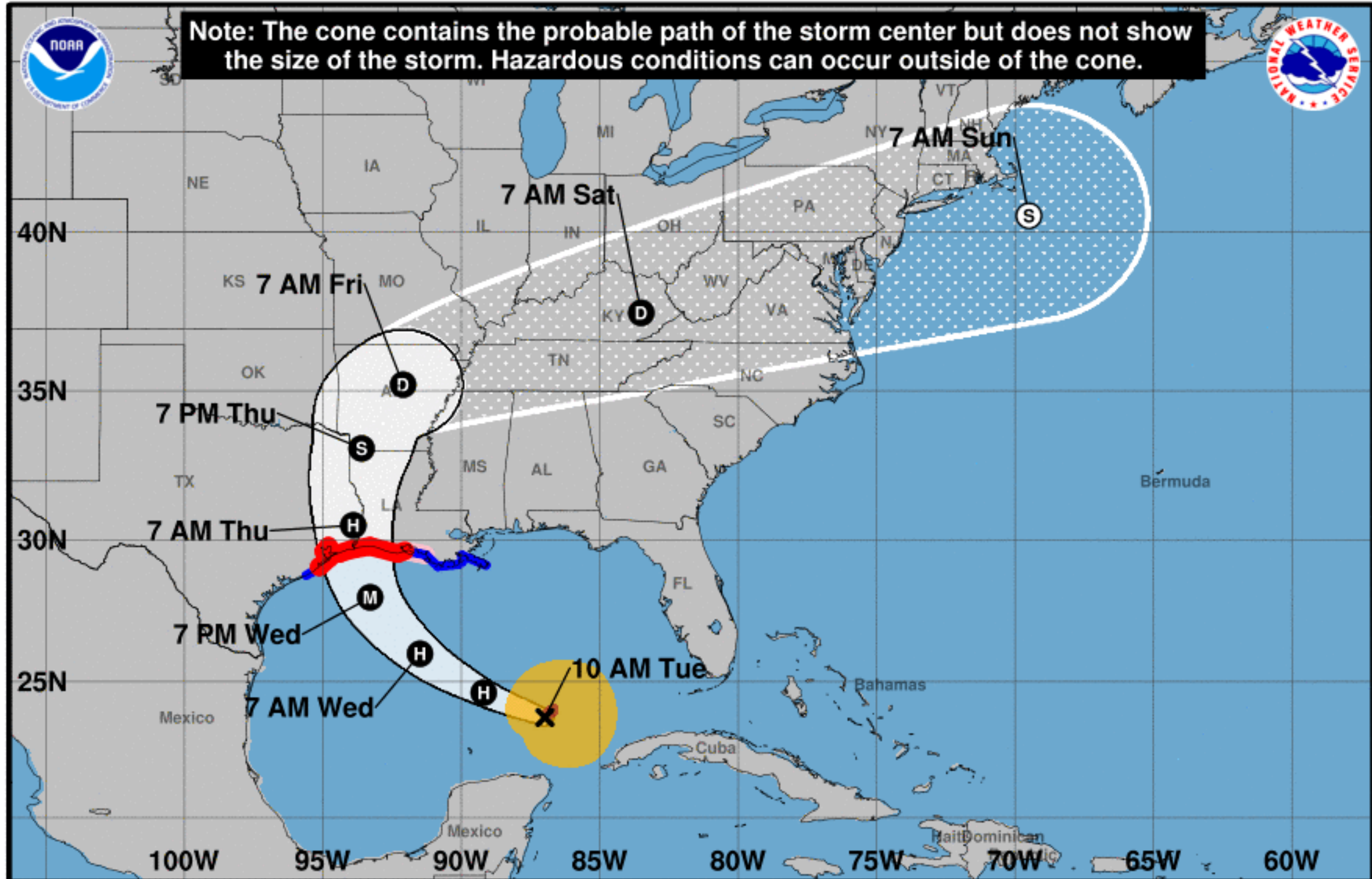
Watches: Hurricane (pink) Trop Stm (yellow)

Warnings: Hurricane (red) Trop Stm (blue)

Current wind extent: Hurricane (brown) Trop Stm (orange)

What is being visualized?

What are the strengths and weaknesses of this visualization?

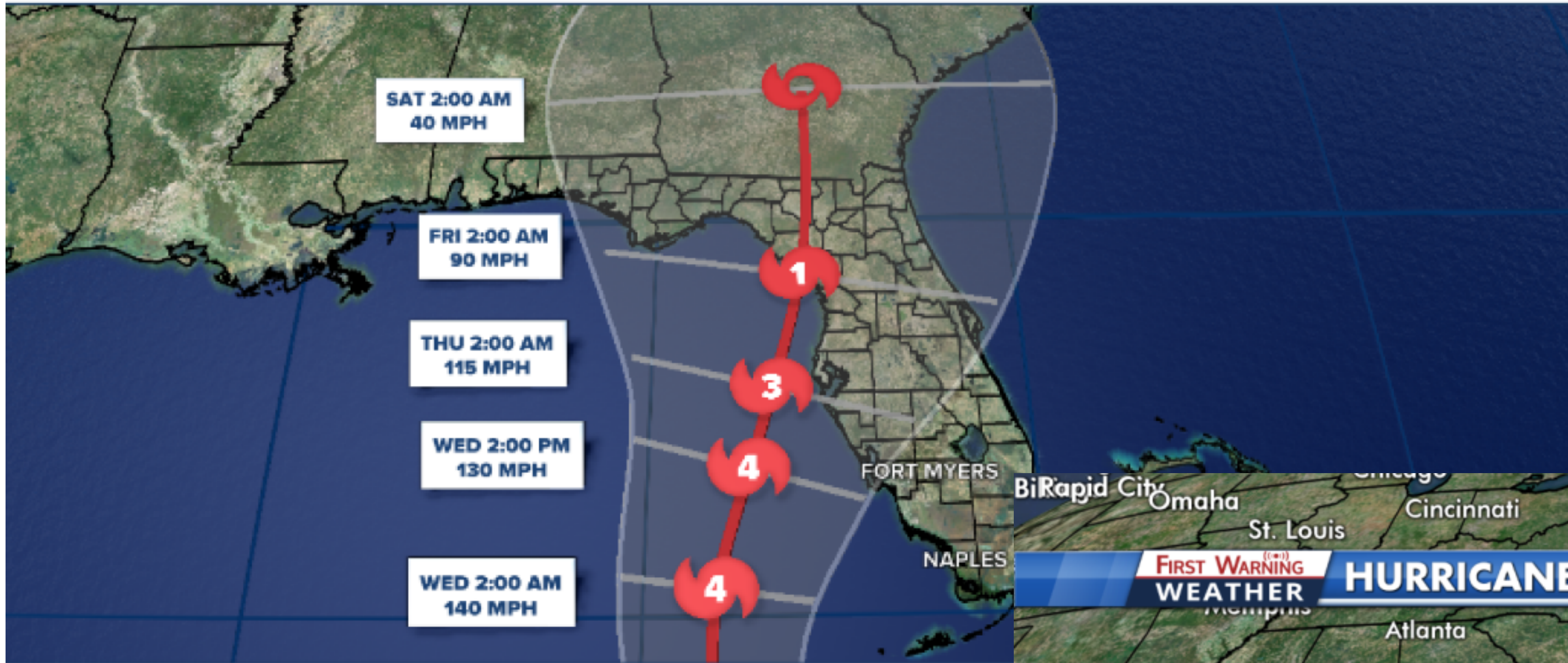


What is being visualized?

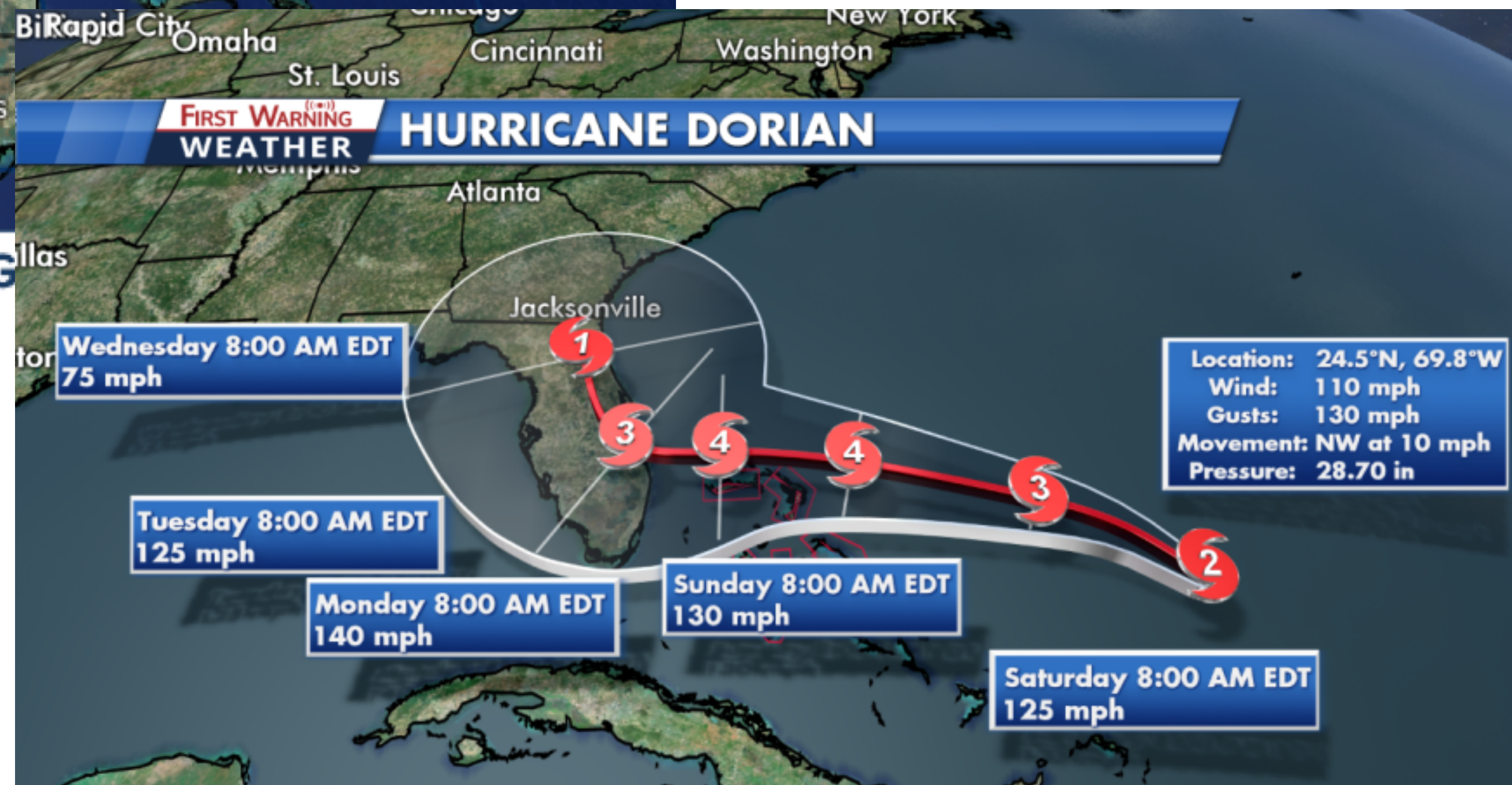
What are the strengths and weaknesses of this visualization?

FOX 4 HURRICANE IAN

5:00 AM ADVISORY

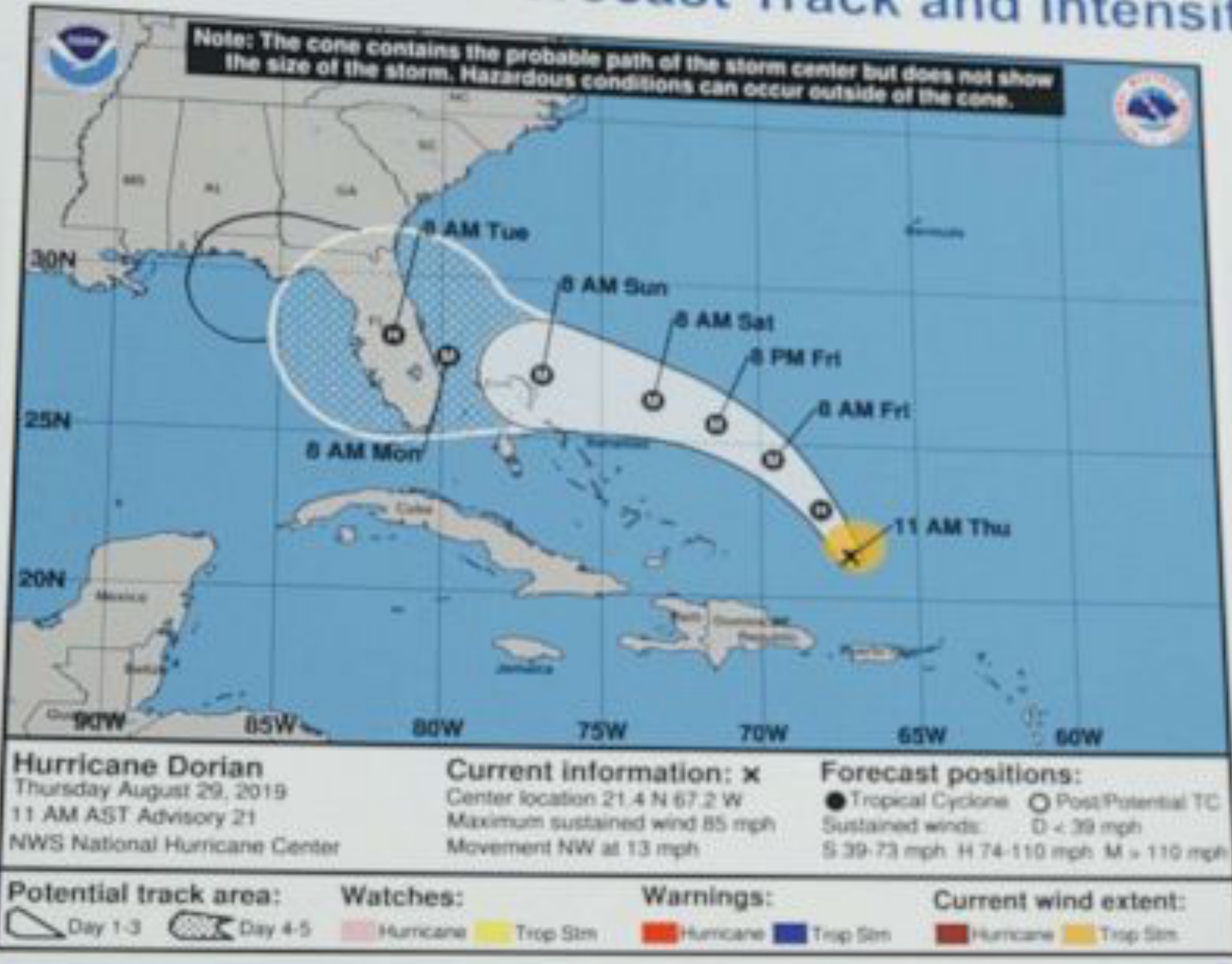


CURRENT LOCATION: 90 MI SW OF G





Hurricane Dorian Forecast Track and Intensity

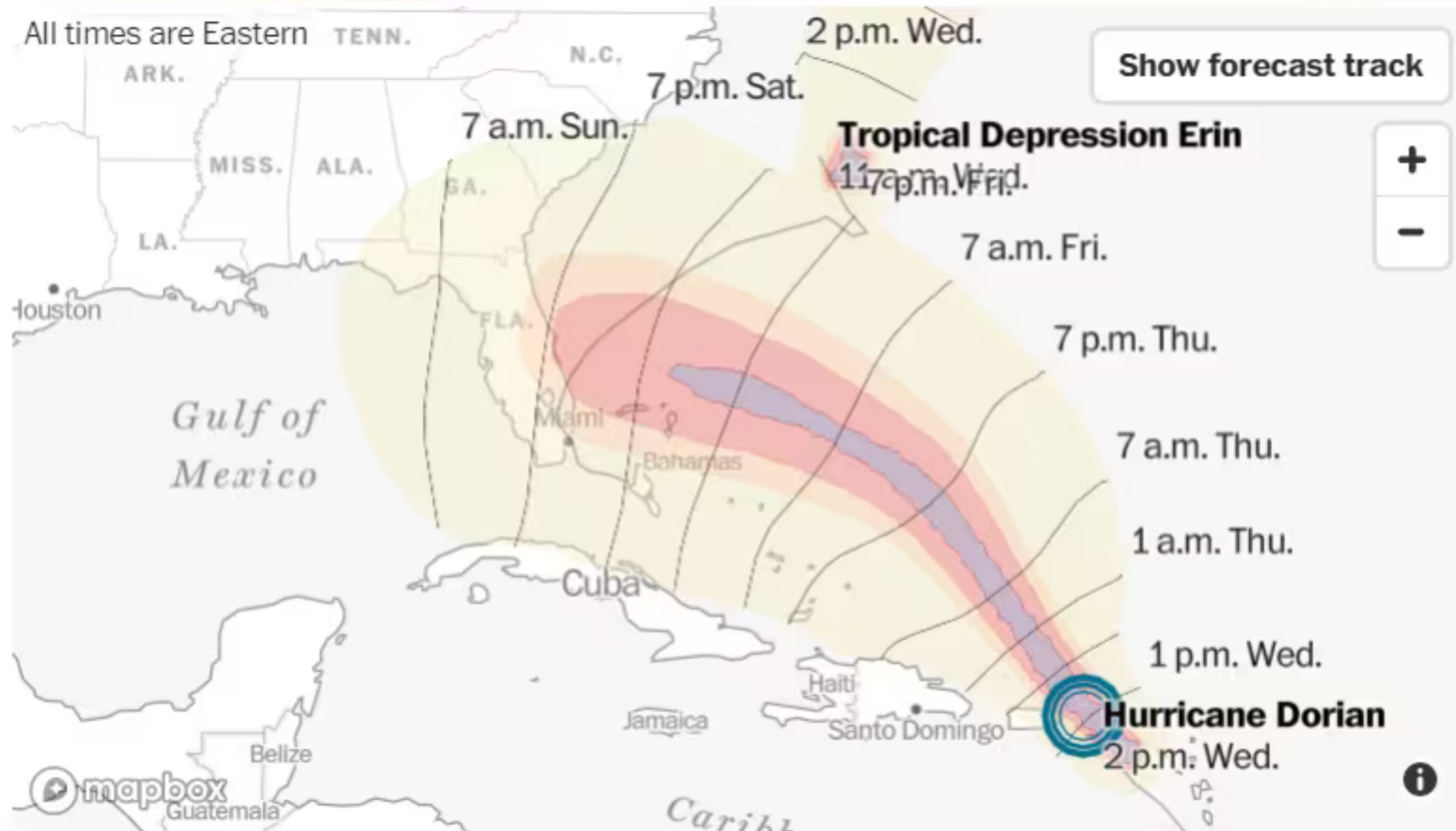


Five-day chance of tropical-storm-force winds



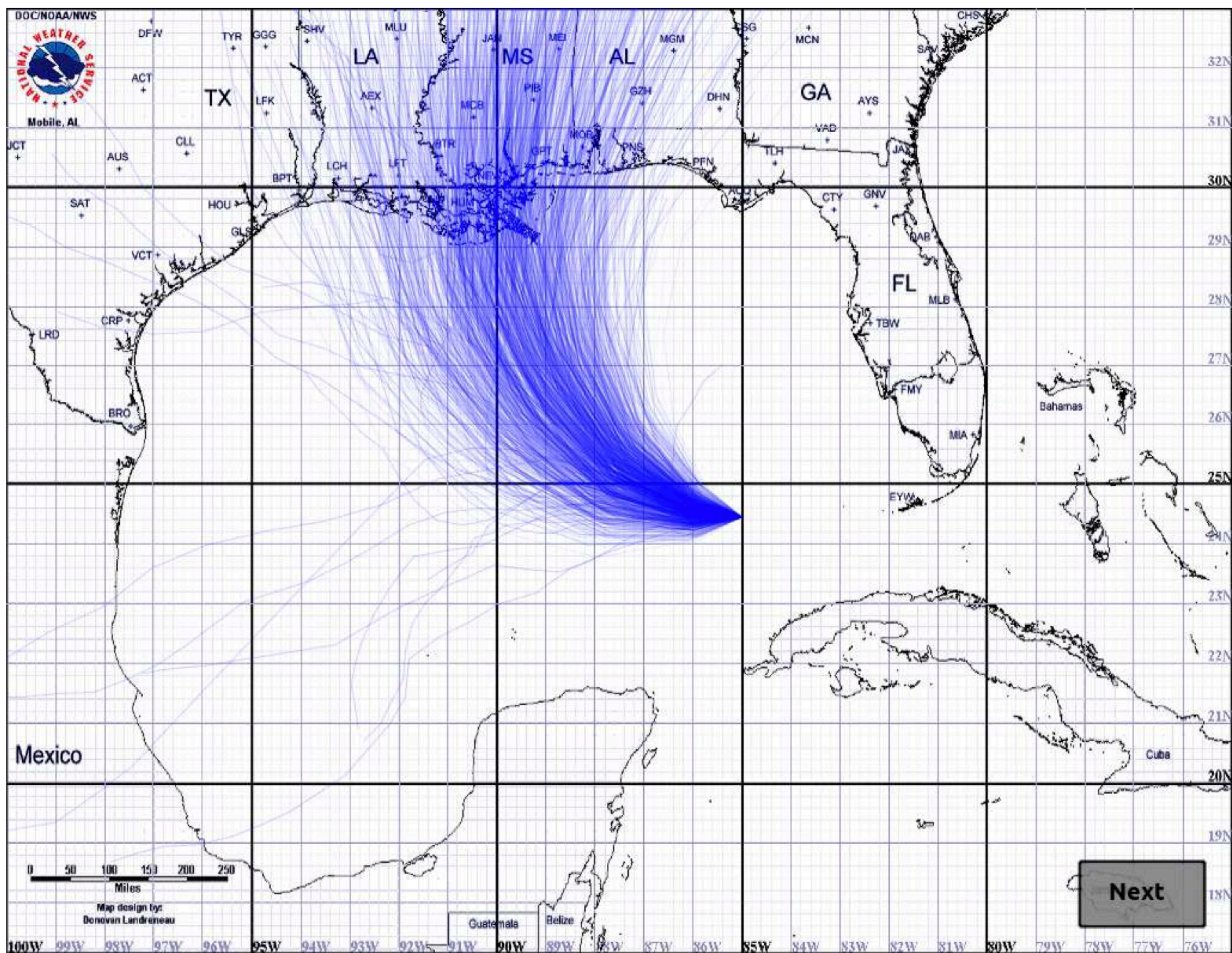
Current extent of tropical-storm-force winds

Major hurricane (>110 mph)
 Hurricane (74-110 mph)
 Tropical storm (39-73 mph)
 Tropical depression (<39 mph)



Source: National Weather Service. Note: Impact lines represent the earliest reasonable arrival time of tropical-storm-force winds.

[Cox et al., 2013]

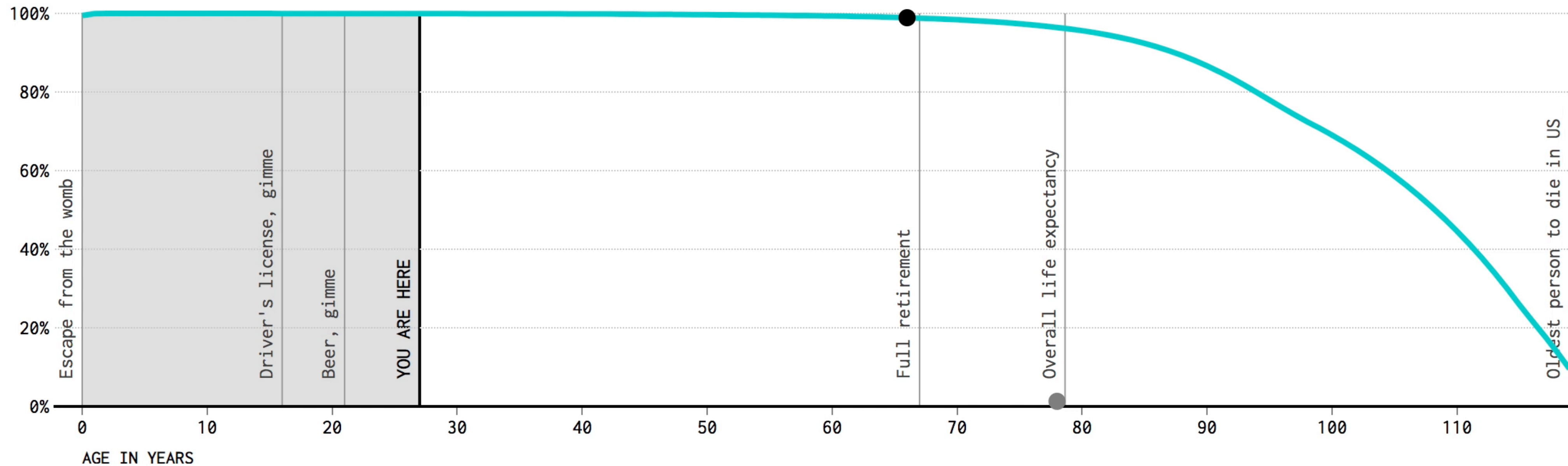


Visualizing Uncertainty: Set of "Draws"

I am **female** and currently **27** years old.

SLOW
FAST

PROBABILITY OF LIVING TO NEXT YEAR



Probabilities For Years Left to Live

0 to 9	10 to 19	20 to 29	30 to 39	40 to 49	50 or more
0%	0%	0%	9%	14%	77%
(0)	(0)	(0)	(2)	(3)	(17)

[Yau, 2015]

Visualizing Uncertainty: Set of “Draws”

The New York Times

• TheUpshot

STATISTICAL NOISE

How Not to Be Misled by the Jobs Report

If the economy actually added 150,000 jobs last month, it would be possible to see any of these headlines:

The jobs number is just an estimate, and it comes with uncertainty.



Visualizing Uncertainty: Set of "Draws"

The New York Times

TheUpshot

STATISTICAL NOISE

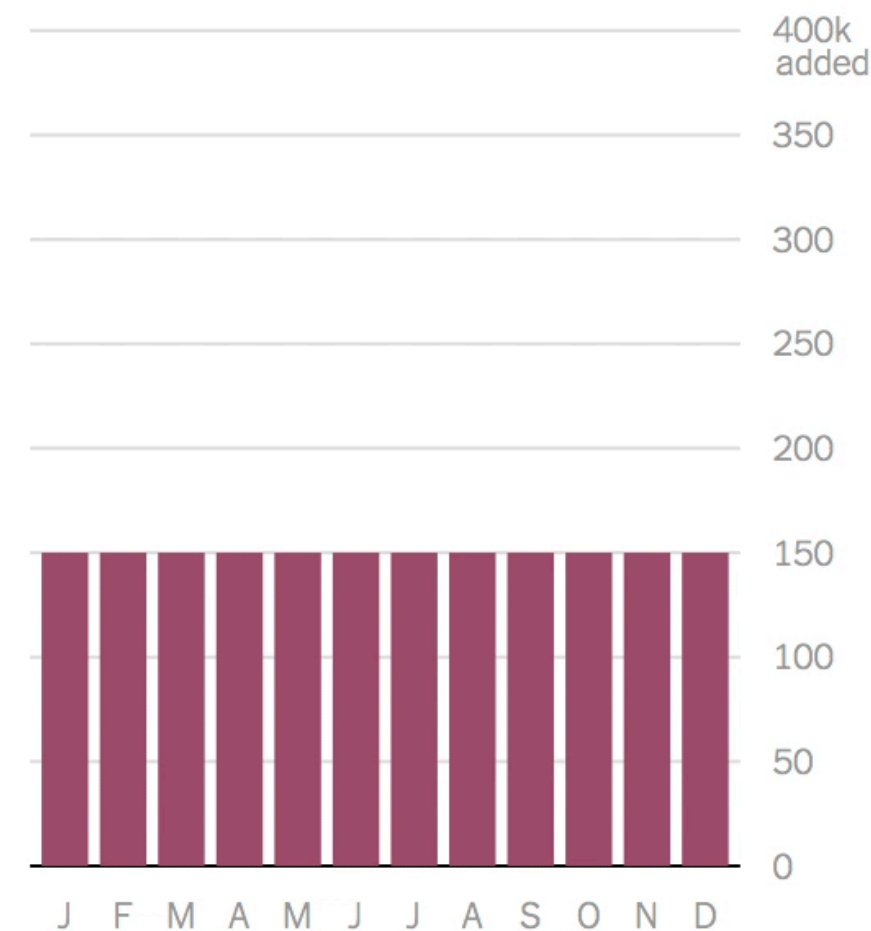
How Not to Be Misled by the Jobs Report

If the economy actually added 150,000 jobs last month, it would be possible to see any of these headlines:

The jobs number is just an estimate, and it comes with uncertainty.

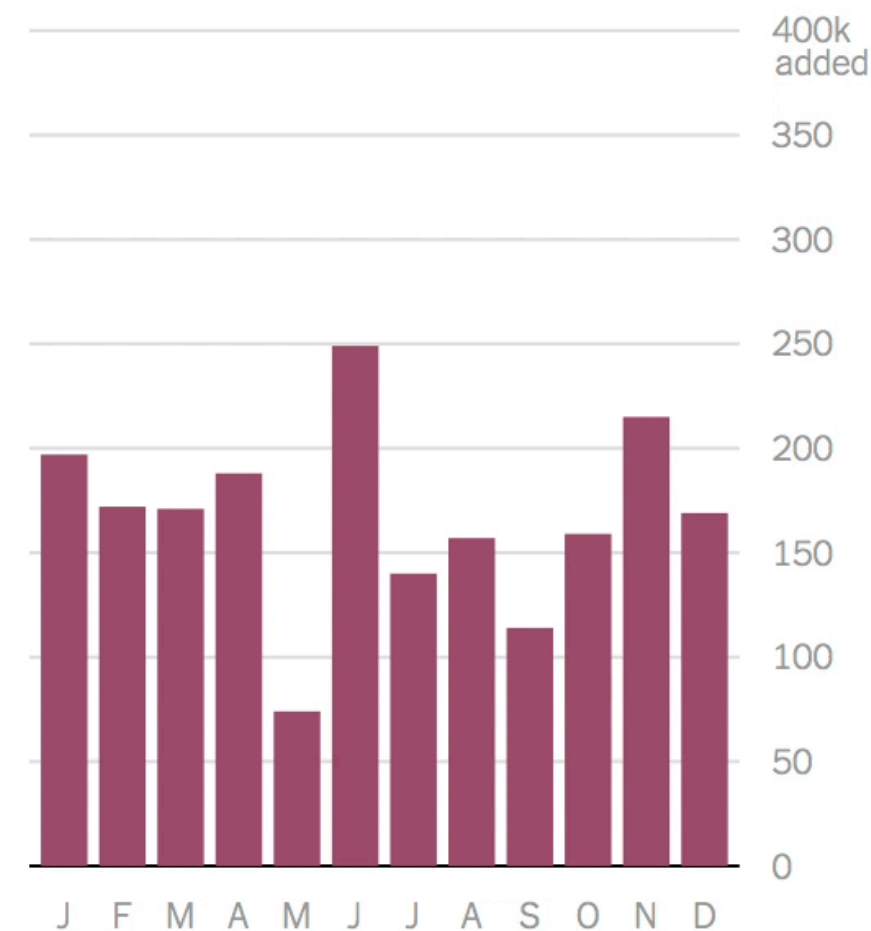


If job growth **were actually steady** over the last 12 months...

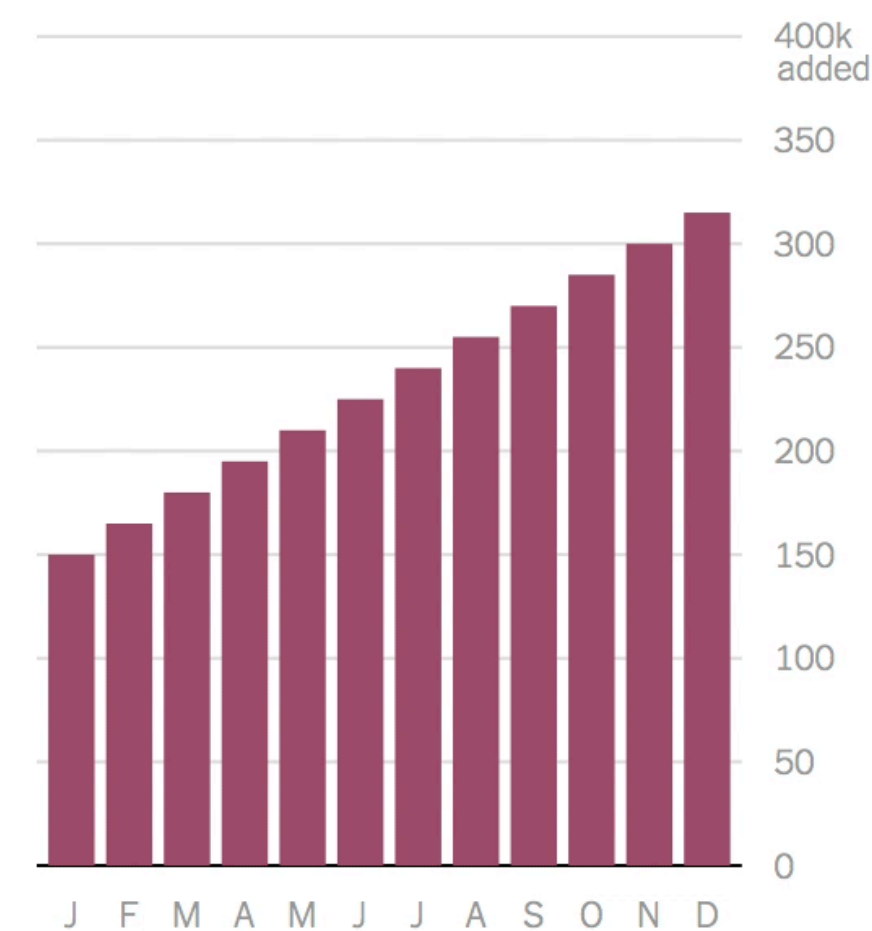


...the jobs report **could look like this:**

[Play](#)

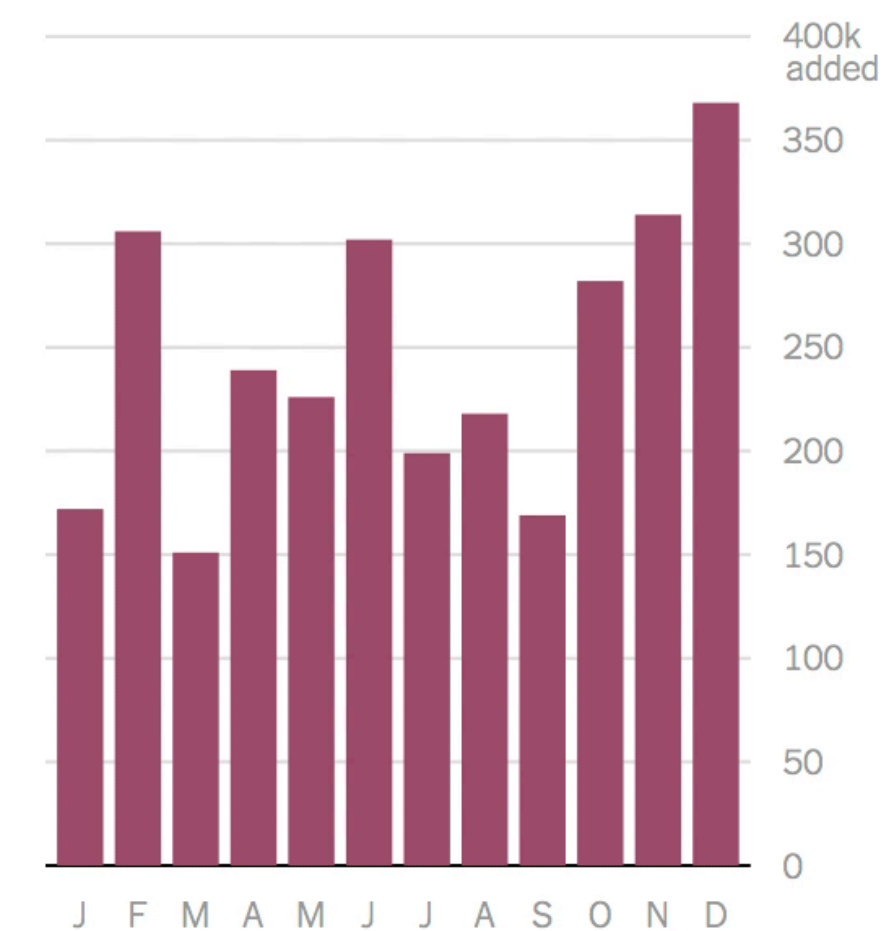


If job growth **had been accelerating...**

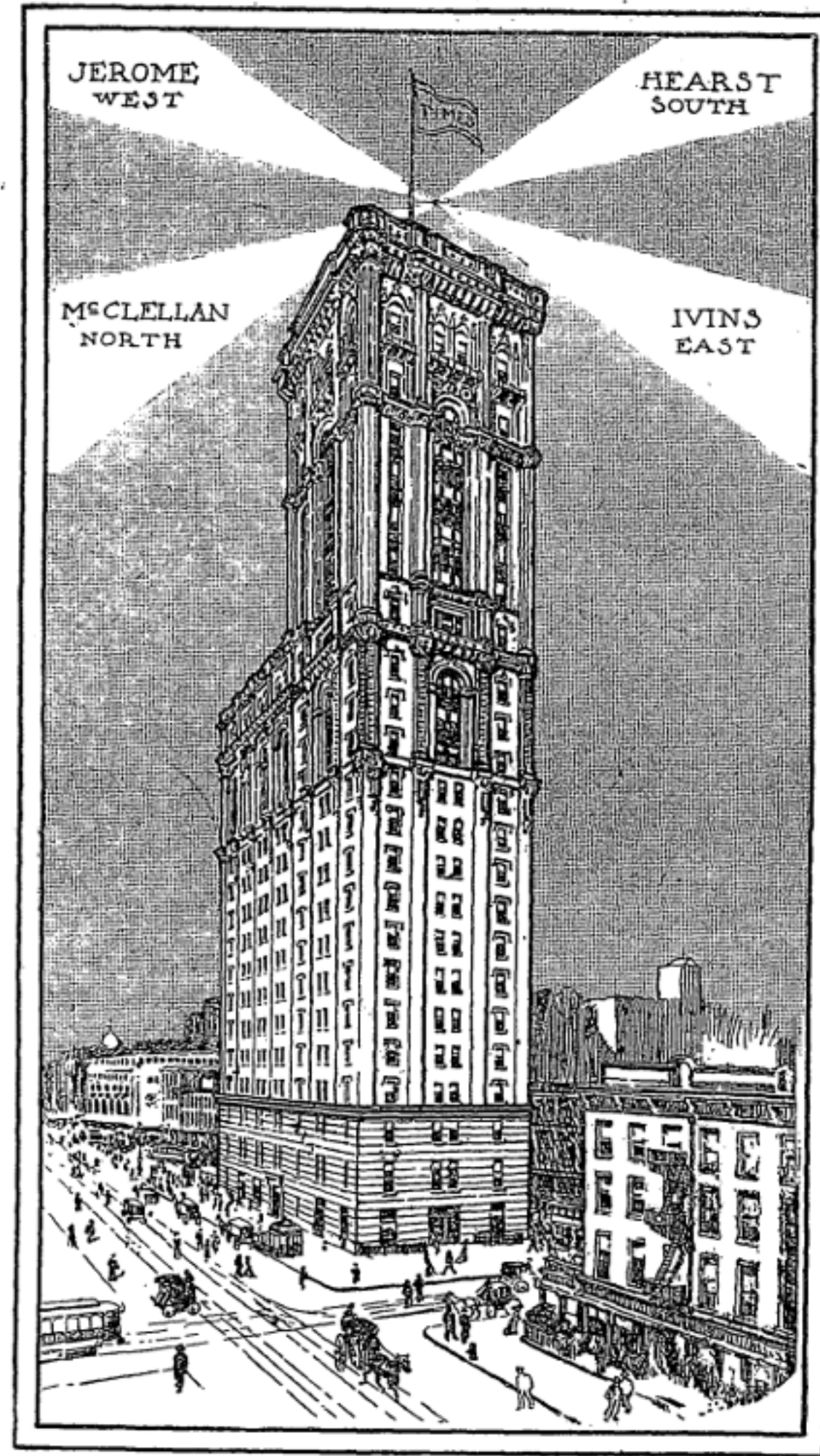


...the jobs report **could look like this:**

[Play](#)



ELECTION RESULTS BY SEARCHLIGHT.



The Times Election Searchlight Code.

News Will Be Flashed from the Tower of The Times Building on Tuesday Night.

The results of the election next Tuesday night will be flashed by electric light from the tower of the Times Building, so that for miles around people will be able to tell which of the candidates has won.

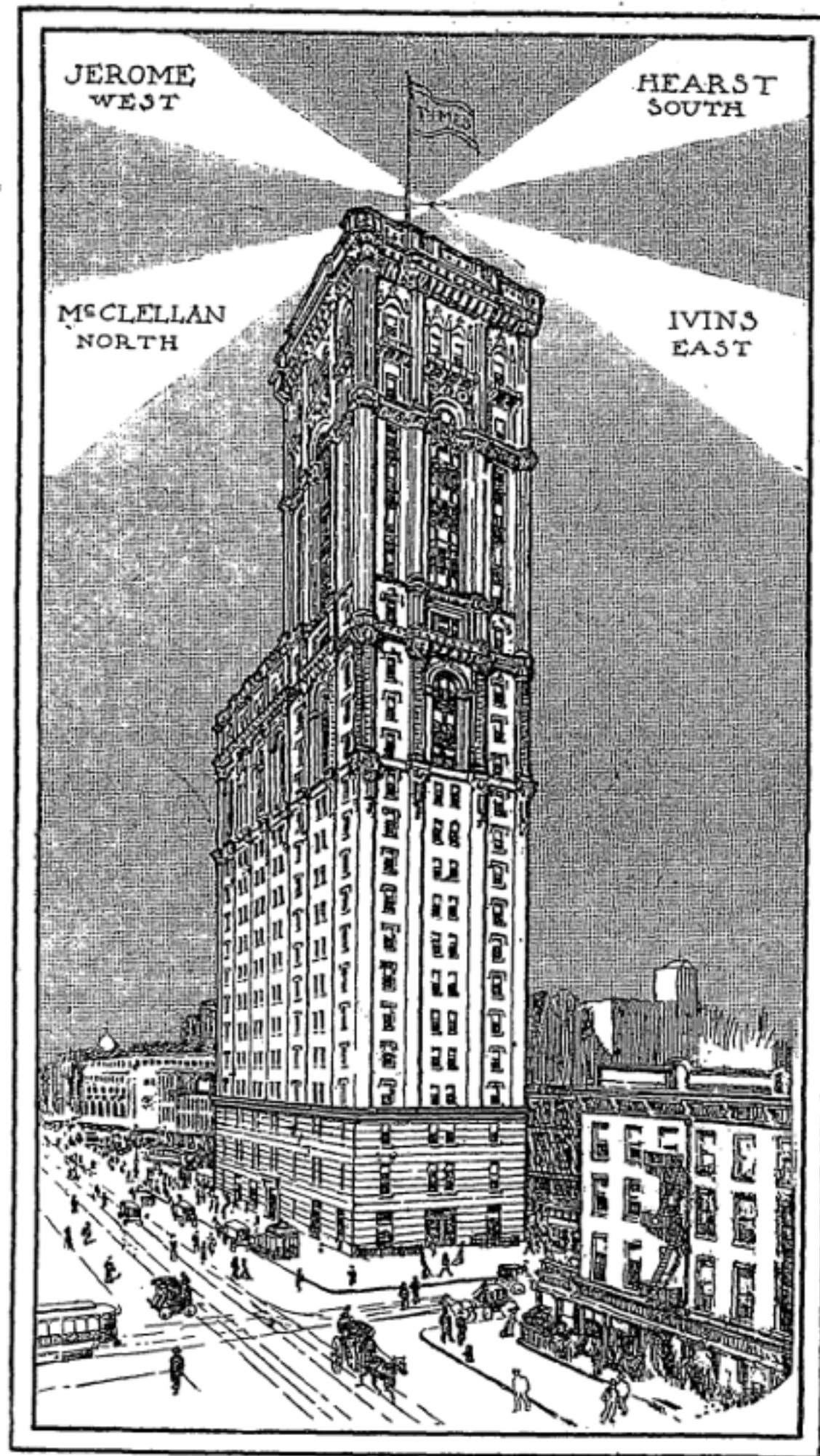
This will be entirely separate and distinct from the elaborate bulletin service which THE TIMES will also maintain. To display the detailed bulletins so that the crowds can see them easily and comfortably, a stereopticon machine will be set up in the triangle north of the Times Building and the bulletins displayed on canvas stretched from the north side of the building. There will be a similar

service at the Harlem office of THE TIMES, 129 West 125th Street.

The electric signals from the tower of the Times Building will be flashed from a point 365 feet above the street level. A steady light to the north will show that McClellan has been elected; a steady light to the east will indicate Ivins's election, and a steady light to the south will indicate that Hearst has won.

Jerome's election will be indicated by a steady light to the west. A light to the north, waving from east to west, will indicate Osborne's election. A light to the south, waving from east to west, will indicate Shearn's election.

ELECTION RESULTS BY SEARCHLIGHT.



The Times Election Searchlight Code.

News Will Be Flashed from the Tower of The Times Building on Tuesday Night.

The results of the election next Tuesday night will be flashed by electric light from the tower of the Times Building, so that for miles around people will be able to tell which of the candidates has won.

This will be entirely separate and distinct from the elaborate bulletin service which THE TIMES will also maintain. To display the detailed bulletins so that the crowds can see them easily and comfortably, a stereopticon machine will be set up in the triangle north of the Times Building and the bulletins displayed on canvas stretched from the north side of the building. There will be a similar

service at the Harlem office of THE TIMES, 129 West 125th Street.

The electric signals from the tower of the Times Building will be flashed from a point 365 feet above the street level. A steady light to the north will show that McClellan has been elected; a steady light to the east will indicate Ivins's election, and a steady light to the south will indicate that Hearst has won.

Jerome's election will be indicated by a steady light to the west. A light to the north, waving from east to west, will indicate Osborne's election. A light to the south, waving from east to west, will indicate Shearn's election.

Election Bulletins

BY BOMBS.

TUESDAY NIGHT

THE TRIBUNE

will send up from the roof of the

GREAT NORTHERN HOTEL

hourly, shells containing blue and red stars—exactly on the hour—at 7, 8, 9, 10, 11 p. m. 12 midnight, 1 and 2 a. m. Wednesday morning, unless election is decided earlier, in which case twelve bombs will be sent up in rapid succession. Blue to indicate McKinley's election. Red to indicate Bryan's election.

SIX BOMBS EVERY HOUR.

The first bomb sent up, if blue, indicates the returns in **COOK COUNTY** at that hour are favorable to McKinley; if red, favorable to Bryan.

After sixty seconds two bombs will be sent up in rapid succession, and will indicate, if blue, that returns from **ILLINOIS** favor McKinley; if red, Bryan.

After sixty seconds more three bombs will be sent up in rapid succession, and if blue will indicate that at that hour returns from the **entire country** favor McKinley; if red, Bryan. Each bomb bursts high in the air, scattering a shower of stars.

PART THREE

Visualizing Uncertainty

Visualization, EVA, & Uncertainty

A Crash Course

Arvind Satyanarayan

LES VARIABLES DE SÉPARATION DES IMAGES

GRAIN



COULEUR



ORIENTATION



FORME

