National Aeronautics and Space Administration

Telling Meaningful Stories with Data

Big Data Day LA 2018

August 11, 2018

Alyssa Columbus
Datanaut, NASA - National Aeronautics and Space Administration

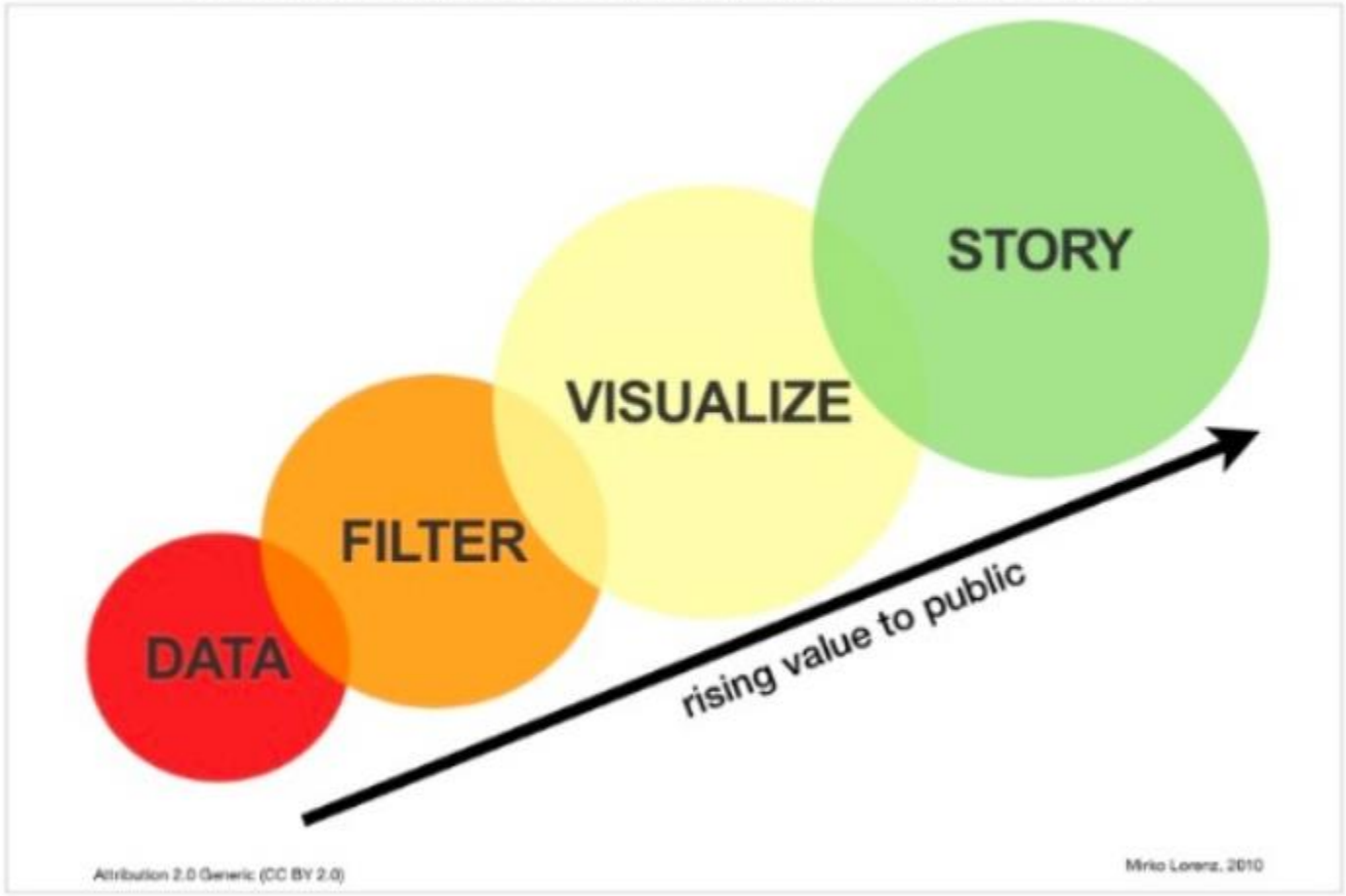
Principles of good

DATA JOURNALISM

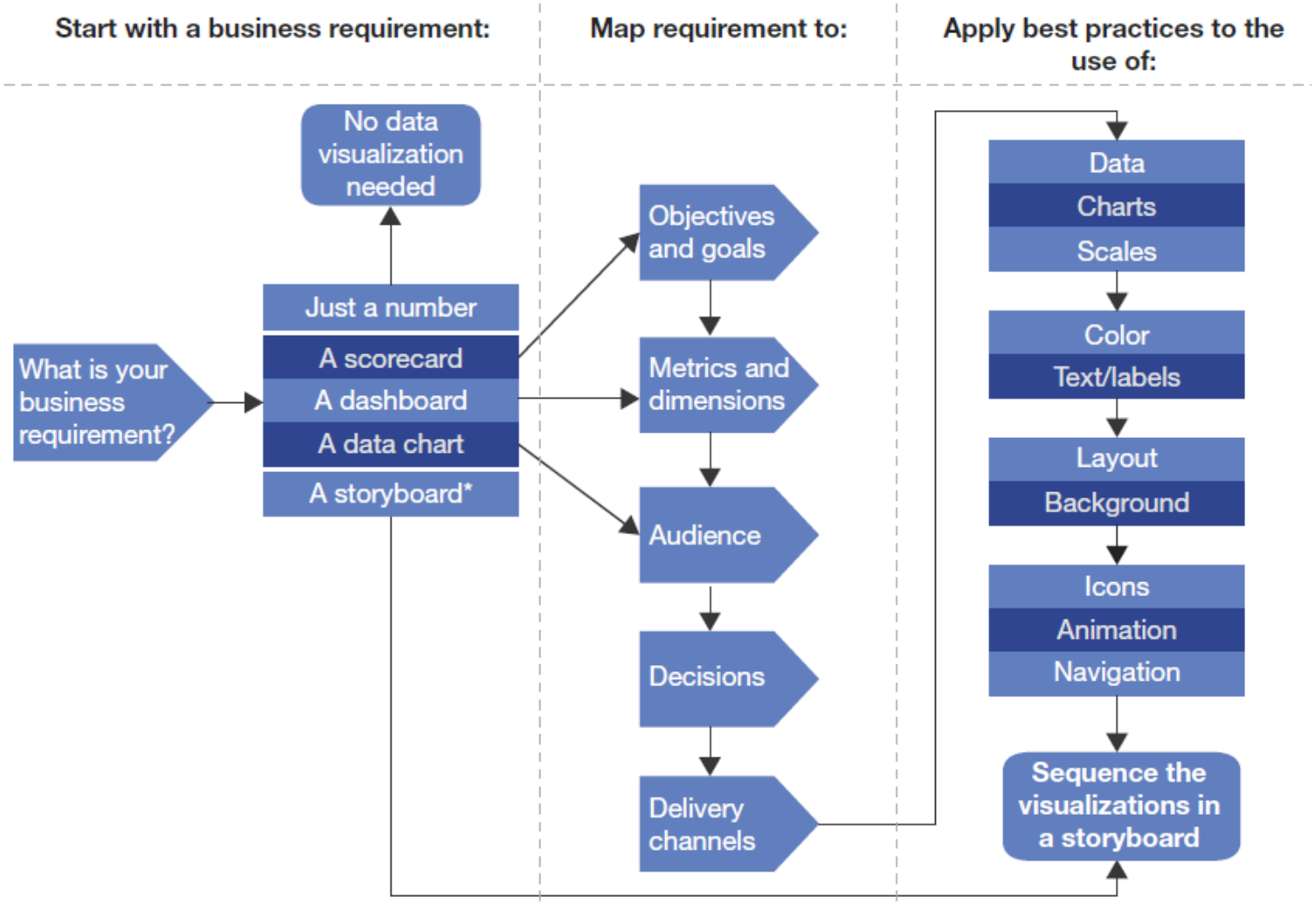
are principles of good

DATA STORYTELLING.

DATA-DRIVEN JOURNALISM = PROCESS



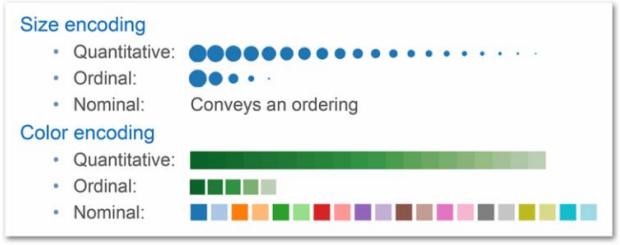
Data Visualization Design Process



*Note: This assumes the storyboard will leverage existing charts, dashboards, and scoreboards.

Source: Forrester, Improve Your Storytelling With Data Visualization Best Practices December 28, 2017

Visualization Techniques



Source: Jock Mackinlay, "The Zen of Visual Analytics," Tableau Software

The Level Of Human Perception Of Each Visualization Technique

Ease of human perception	Visualization type
Easiest	Position along common scale
	Position along nonaligned scales
	Length, height
	Angle, slope
	Area
	Volume
Hardest	Color, density

Matching Analysis Type To The Appropriate Data Visualizations

Analysis type	Analysis variation	Appropriate data visualization
Comparison	Among items	Bar charts
	Over time	Line charts
Distribution	Single variable	Histograms
	Multiple variables	Scatter plots
Composition	Static	Pie charts
	Changing over time	Stacked column and stacked area charts
Relationship	Two variables	Scatter plots
	Three variables	Bubble charts

Example	Encoding	Ordered	Useful Values	Quantitative	Ordinal	Categorical	Relational
	Position, Placement	Yes	Infinite	Good	Good	Good	Good
1, 2, 3; A, B, C	Text Labels	Optional (alphabetic/ numbered)	Infinite	Good	Good	Good	Good
	Length	Yes	Many	Good	Good		
	Size, Area	Yes	Many	Good	Good		
	Angle	Yes	Medium/ Few	Good	Good		
	Pattern Density	Yes	Few	Good	Good		
	Weight, Boldness	Yes	Few		Good		
	Saturation, Brightness	Yes	Few		Good		
	Color	No	Few (<20)			Good	
	Shape, Icon	No	Medium			Good	
	Pattern, Texture	No	Medium			Good	
	Enclosure, Connection	No	Infinite			Good	Good
	Line Pattern	No	Few				Good
	Line Endings	No	Few				Good
	Line Weight	Yes	Few		Good		

Source: Noah Iliinsky, "Choosing visual properties for successful visualizations," IBM

Selecting The Right Visualizations

Change Over Time

Shows how a measure changes over time, and allows the user to highlight temporal trends

- Line Chart: Highlights potential trends in data
- Bar Chart: Highlights comparison between individual values

Comparison

Shows the comparison of categorical values, where the data does not have any intrinsic order, for example, a list of products

- Bar Chart: used for comparing categorical values
- Trellis: uses multiple views to show different partitions of a dataset

Ranking

Shows the top or bottom N values to emphasize the largest or smallest values

- Bar Chart: shows categorical values in decreasing or increasing order

Part-To-Whole

Shows how the categories contribute to the whole value

- Bar Chart: set to % scale
- Pie Chart: Compares percentage values
- Stacked Bar Chart: shows overall measure total

Distribution

Shows how a measure is spread across its domain

- Histogram: Column Chart showing the count of binned measure values
- Box Plot: shows distributions for different categorical values
- Heat and Tree Map: shows the distribution of measure values

Correlation

Shows, whether there is a potential correlation between two measures

- Scatter Plot: highlights potential correlation of two measures
- Trellis: uses multiple views to show different partitions of a dataset

Overview

Shows the exact values in table format

- Table: highlights exact values

Geographical Information and Maps

Shows the geographical distribution of measure values

- Choropleth Chart: highlights geographical data by colouring geographical areas according to their measure values
- Geo Bubble Chart: highlights geographical data by showing them as bubbles on a map

Pro Tips

1.

Less is more. Make every pixel and word count.

2.

Avoid decorative use of graphics.

3.

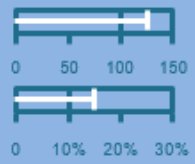
Avoid three-dimensional chart types.

4.

Avoid pie charts.

5.

Start bar charts at zero.



6.

Use bullet graphs instead of gauges to save space.

7.

Use sparklines to show trends on the X-axis.

8.

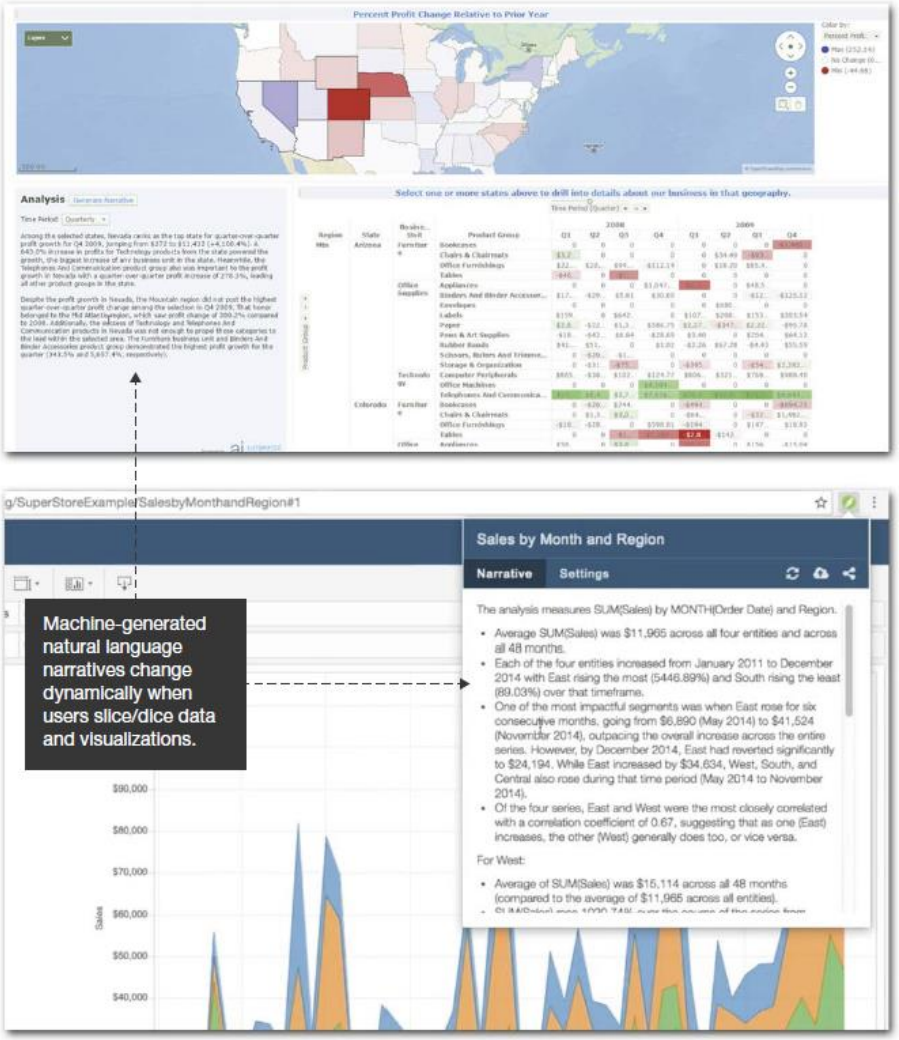
Show time going from left to right on the X-axis.

9.

Use color only to highlight or accentuate meaning.

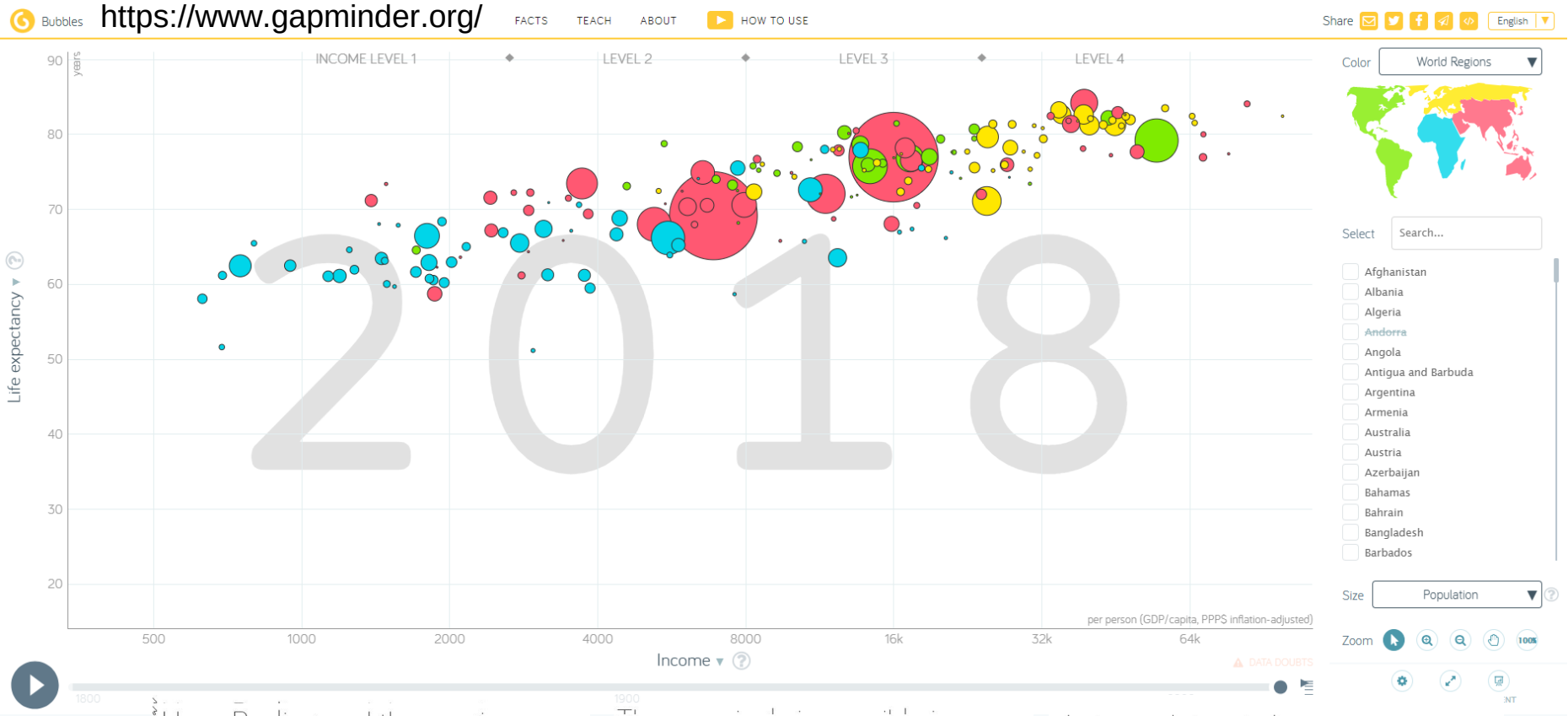
AI-Assisted Analysis

Natural Language Generation Helps Explain Data Visualizations



Source: TIBCO Software (top) and Tableau Software (bottom)

Explore GapMinder



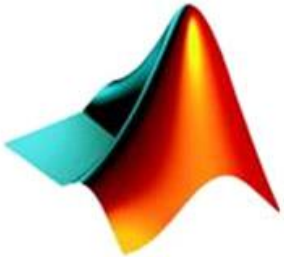
Hans Rosling and the magic washing machine

The seemingly impossible is possible

Let my dataset change your mindset

Tools

Google Developers



MATLAB

WolframAlpha

alteryx

plotly

Data Journalism

DATA LEADS,

BUT STORY TELLS.

Data Journalism

John Snow Cholera Map

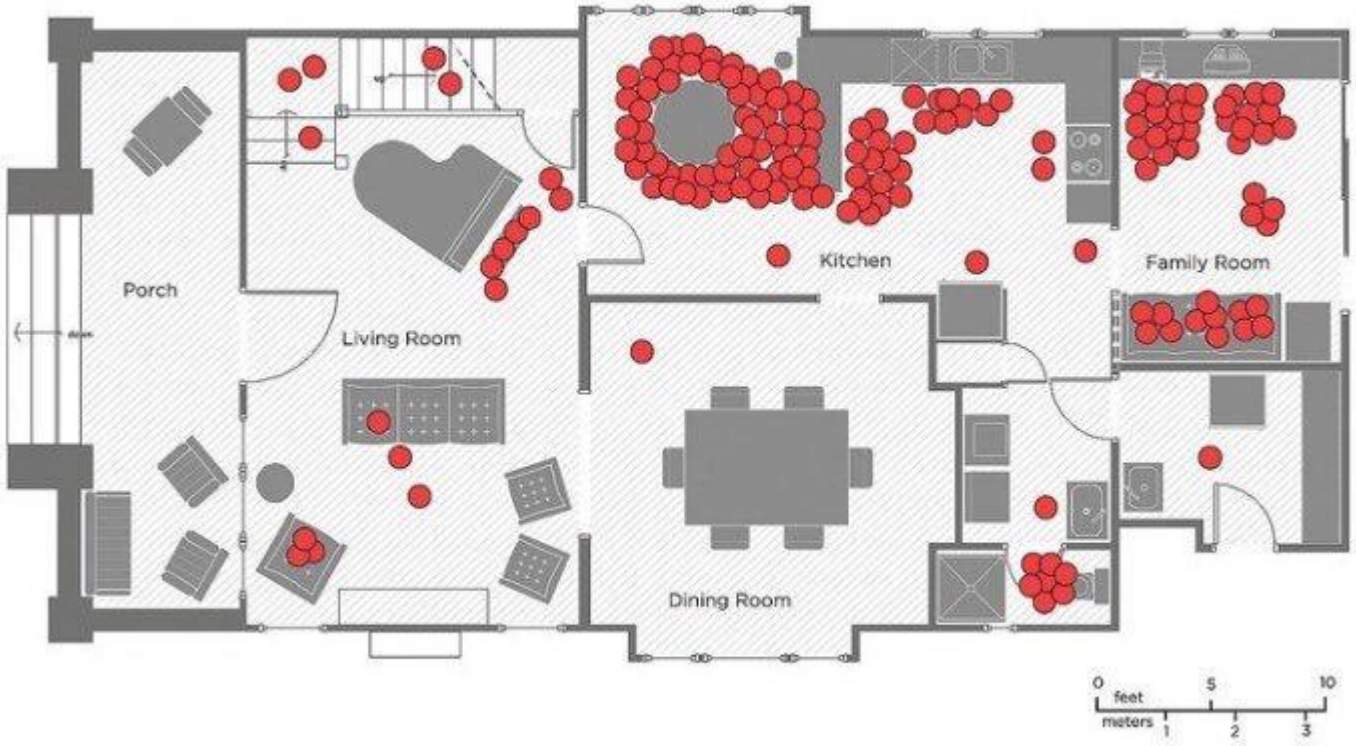
John Snow, 1854

Data Journalism

RIGOROUS,

BUT INTERPRETABLE.

Data Journalism

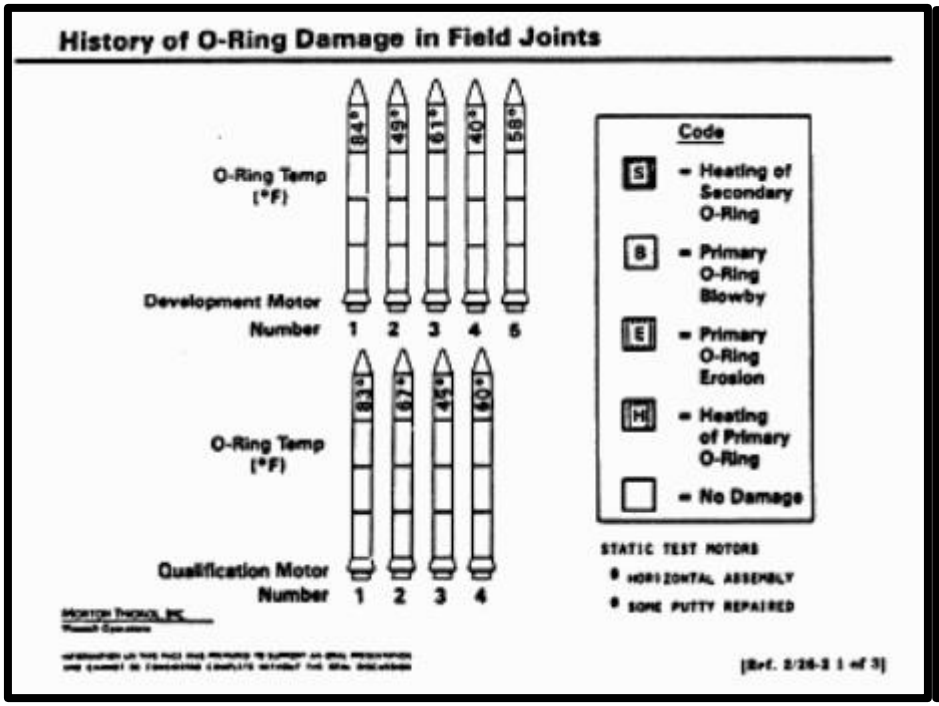


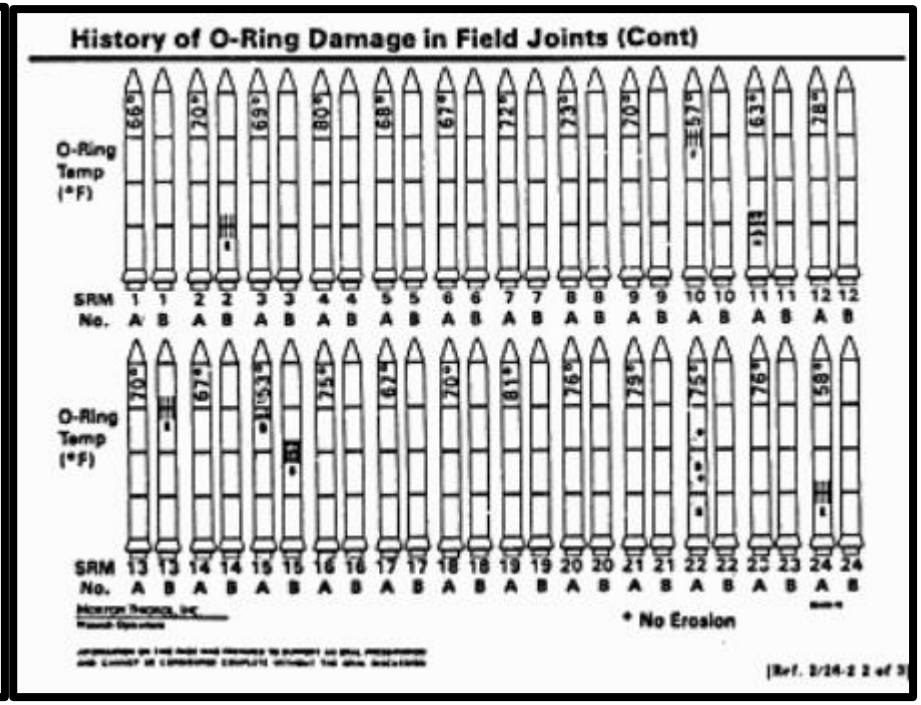
Source: https://legallysociable.com/2018/06/03/americans-can-spend-a-majority-of-their-time-in-a-few-spaces-in-their-home-and-still-want-large-homes/

ACCURATE. FAST.

INDEPENDENT.

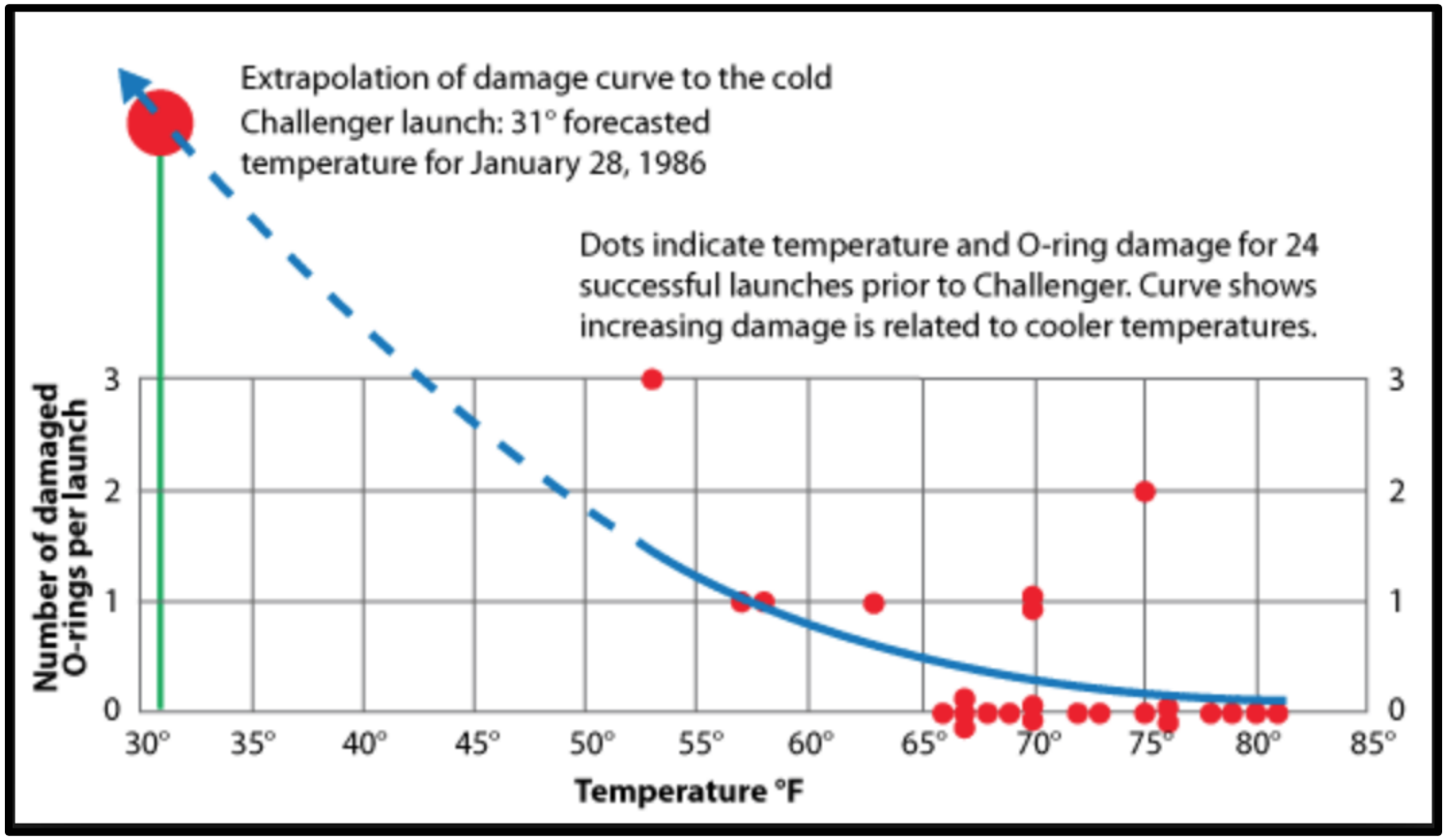
Data Journalism





Source: Presidential Commission on the Space Shuttle Challenger Accident, vol. 5 (Washington, DC: US Government Printing Office, 1986.) pp. 895-896.

Data Journalism



Source: Presidential Commission on the Space Shuttle Challenger Accident, vol. 5 (Washington, DC: US Government Printing Office, 1986.) pp. 895-896.

Data Journalism

NASA

Google News Initiative

https://newsinitiative.withgoogle.com/

Google Public Data Explorer

Google Earth Pro

Tilegrams

Google Crisis Map

Google Fusion Tables

Global Forest Watch

Google Trends

Google Consumer Surveys

Google Data GIF Maker

Flourish

Project Shield

Election Databot

FiveThirtyEight

Good Stories

1. **CHARACTERS**
    - Protagonists & Antagonists
2. **SETTING**
3. **BEGINNING, MIDDLE, END**

Data Stories

1. OUTLIER EXPOSURES

2. WHAT'S TRENDING

3. FORECASTING THE FUTURE

1. OUTLIER EXPOSURES

FiveThirtyEight

MLB Wins Above Replacement - 2018

Season: 2018

ESPN

MLB WAR Leaders»

Players By Season | All-Time Players

Seasonal | Offensive | Defensive | Pitching

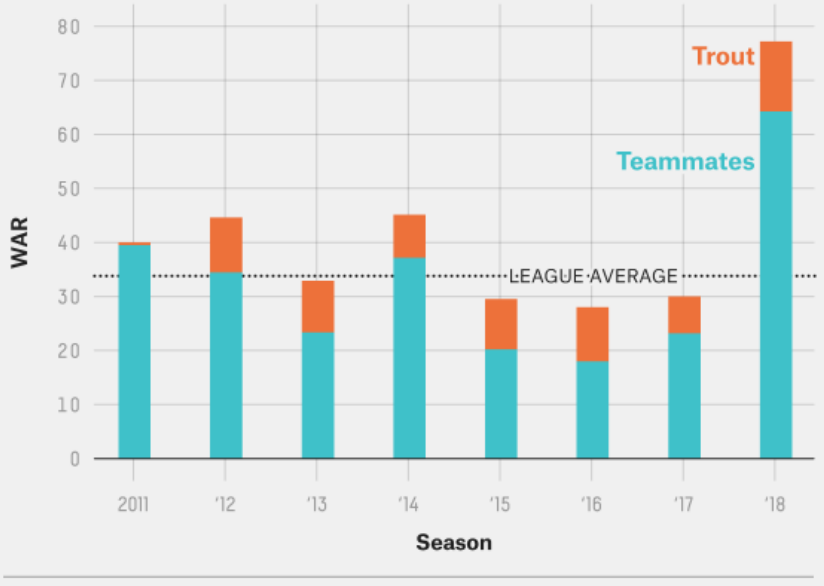
2018 Season WAR Leaders

RK	PLAYER	WAR	OFF	DEF	WAA
1	Mike Trout	7.8	7.2	0.8	6.2
2	Mookie Betts	7.5	6.4	0.8	6.1
3	Jose Ramirez	7.2	6.5	1	5.5
4	Aaron Nola	7	0	0	5.8
5	Jacob deGrom	6.9	0	0	5.7
	Francisco Lindor	6.9	5.7	1.9	5.3
7	Max Scherzer	6.5	0	0	5.2
8	Matt Chapman	6.3	3.5	3.1	4.9
	Chris Sale	6.3	0	0	5
10	Trevor Bauer	5.5	0	0	4.1
	Kyle Freeland	5.5	0	0	4.4

Source: http://www.espn.com/mlb/war/leaders/_/type/seasonal/year/2018

For once, Trout's teammates have been top-notch

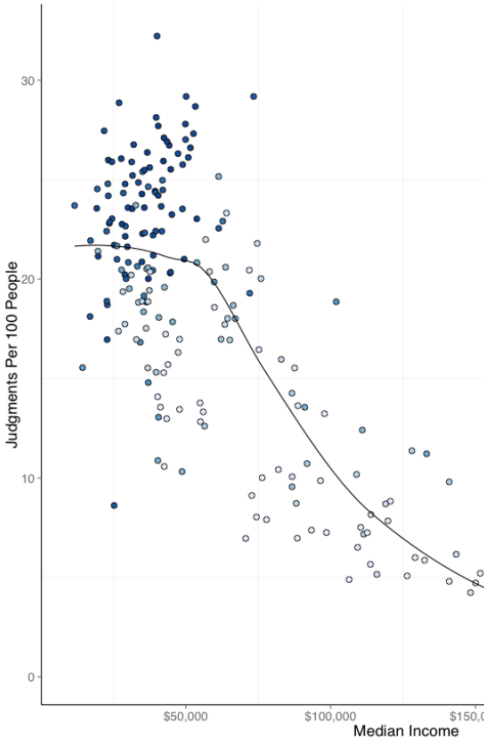
Total wins above replacement (WAR) per 162 games for Mike Trout and the rest of the Los Angeles Angels by season



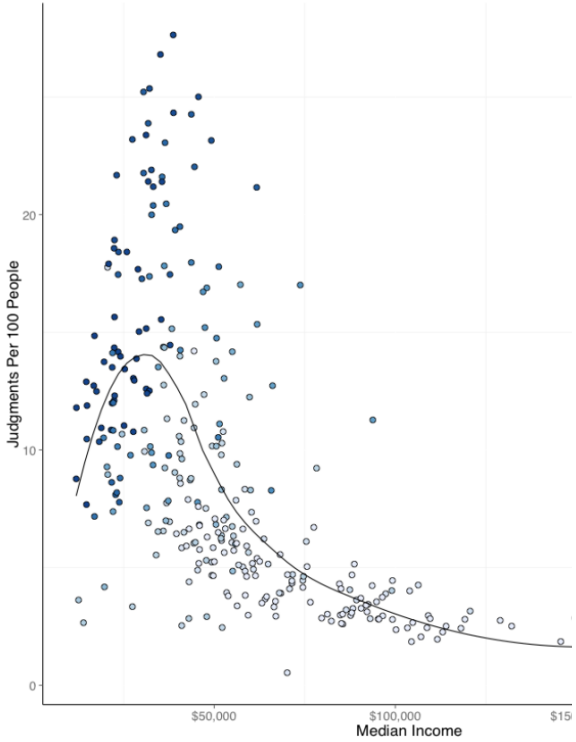
Source: https://fivethirtyeight.com/features/mike-trout-finally-has-some-help/

2. WHAT'S TRENDING

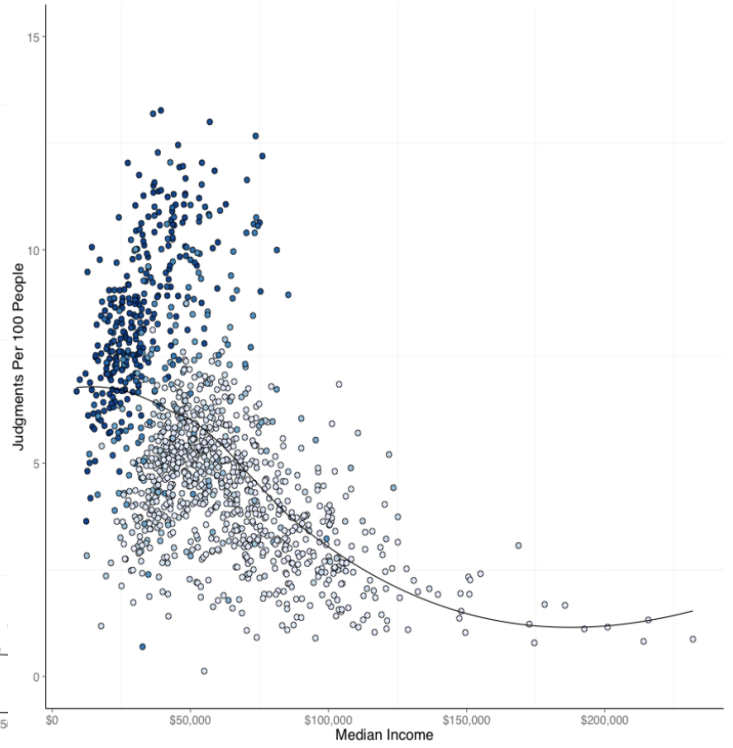
3.2.3. Analysis of Essex County, New Jersey



3.2.1. Analysis of St. Louis County and City, Missouri

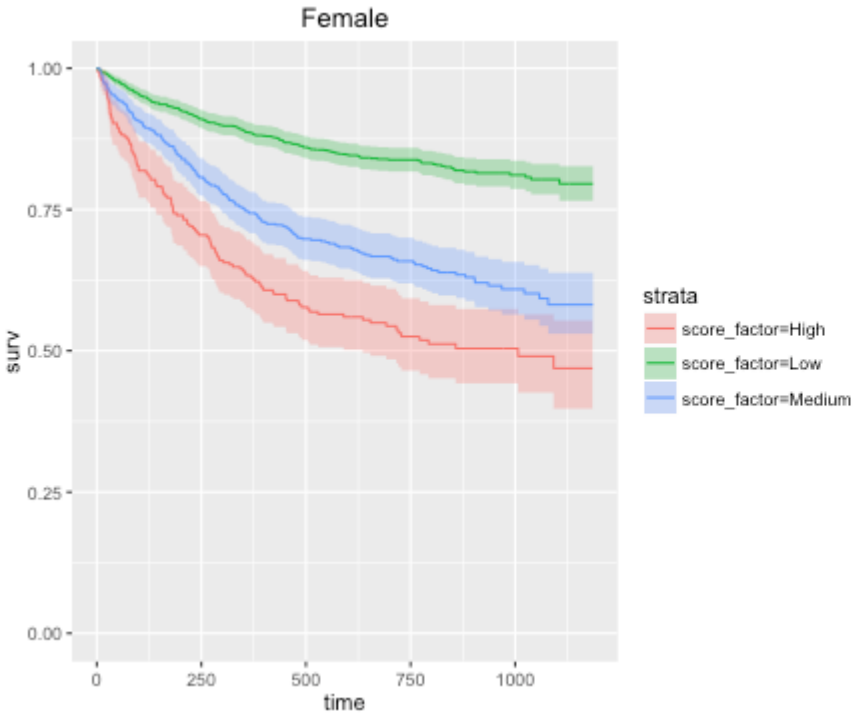


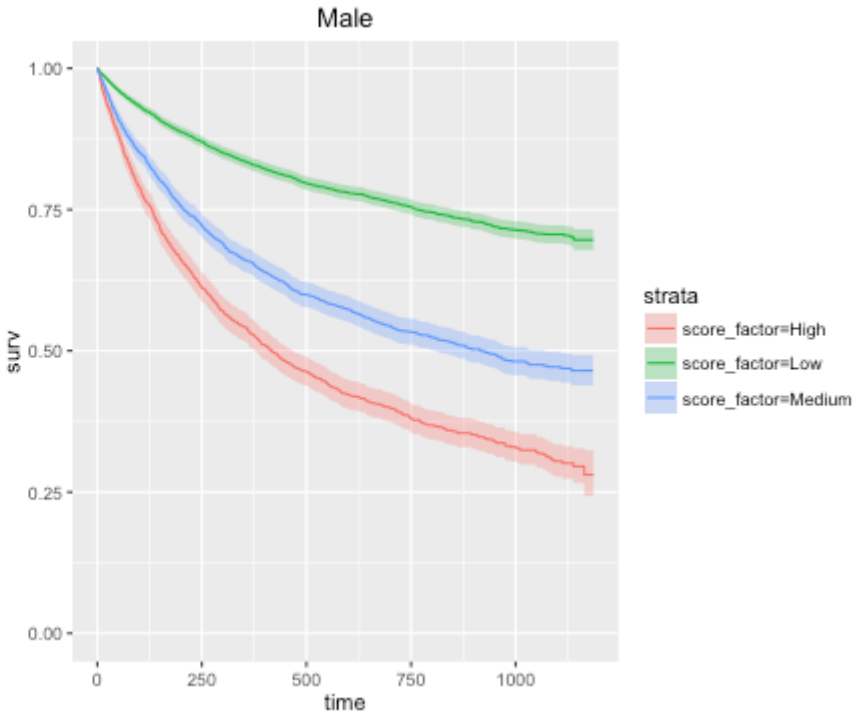
3.2.2. Analysis of Cook County, Illinois



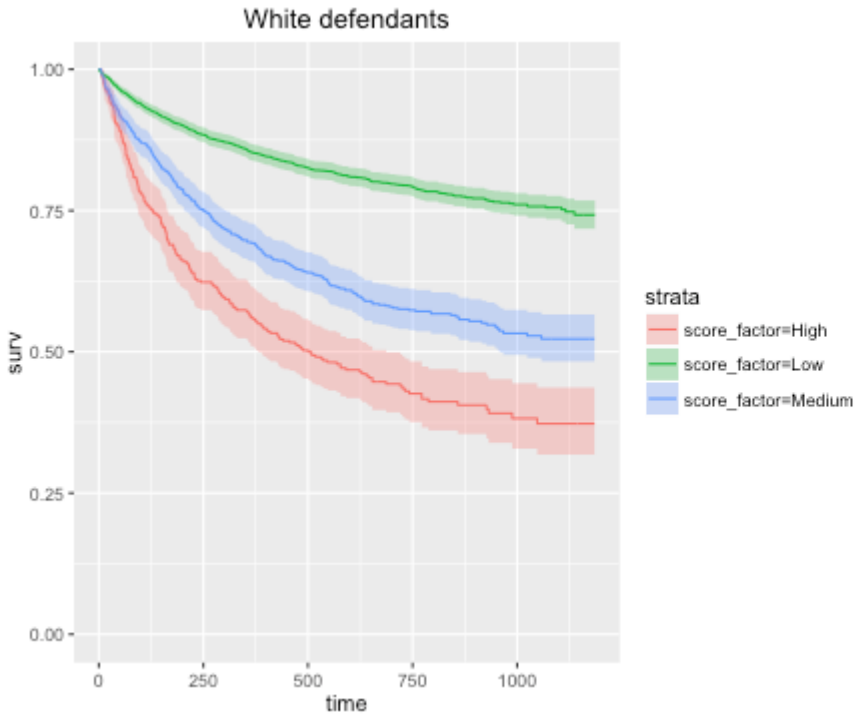
Source: https://www.propublica.org/article/debt-collection-lawsuits-squeeze-black-neighborhoods

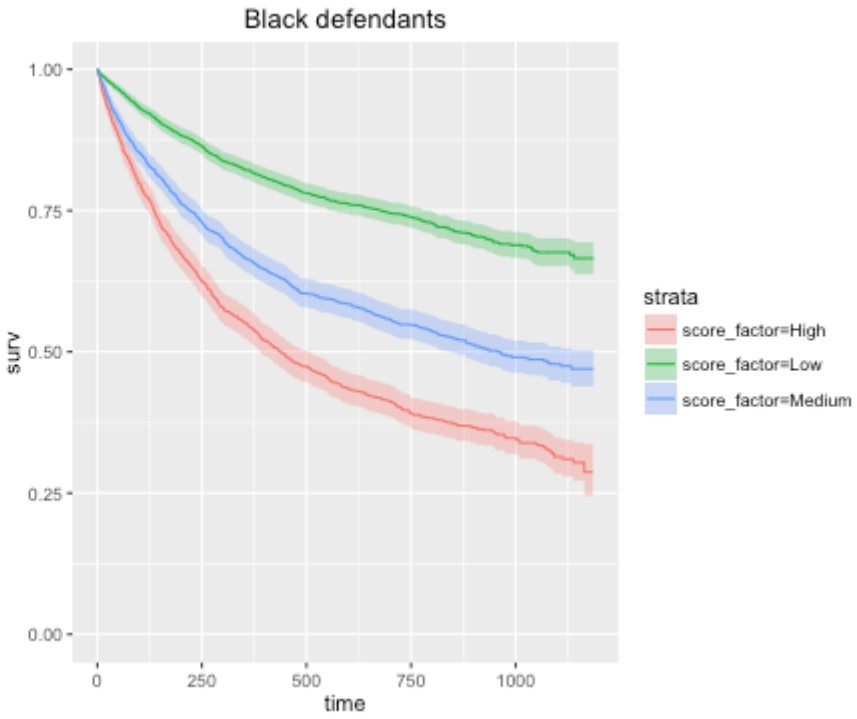
3. FORECASTING THE FUTURE





Source: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

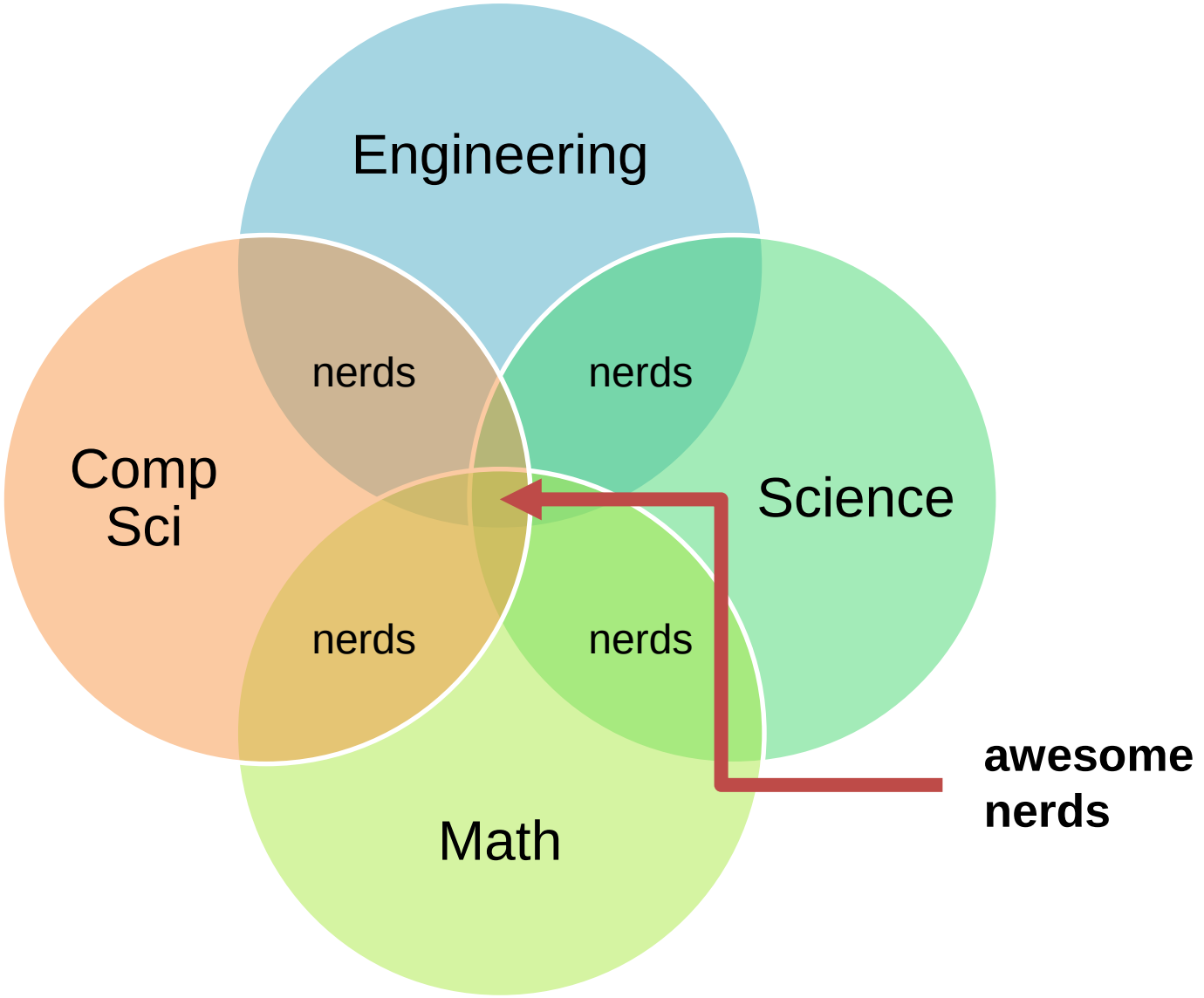




Source: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

Every person has the ability to tell a story.

Data Scientists




story

SUPPLEMENTARY READING

RECOMMENDED READING

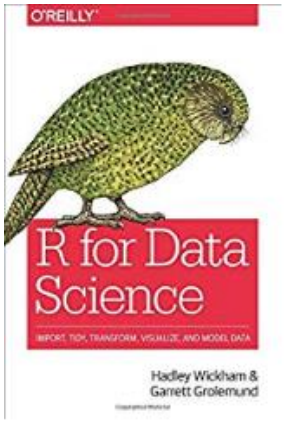
FLOWINGDATA

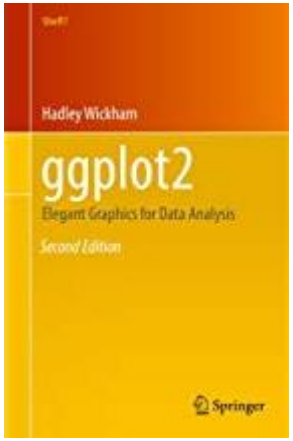
information is beautiful

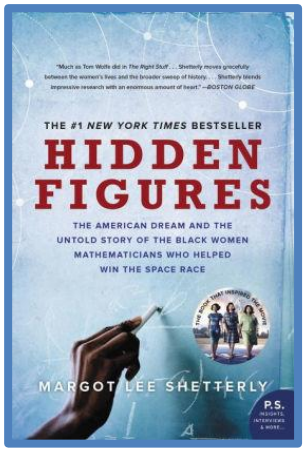
modern data *new generation of data science*

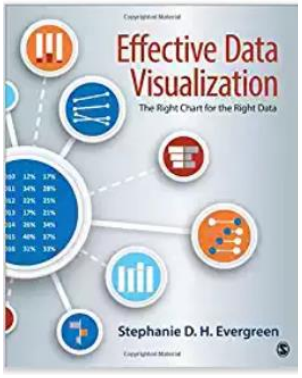
storytelling WITH data

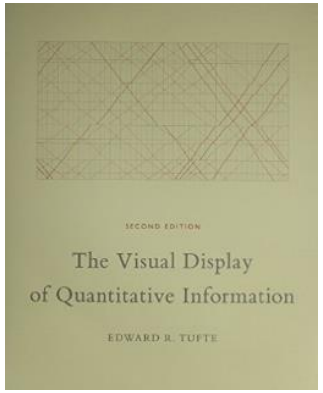
visualising data

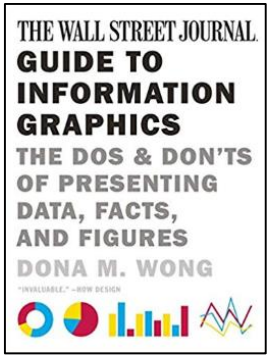


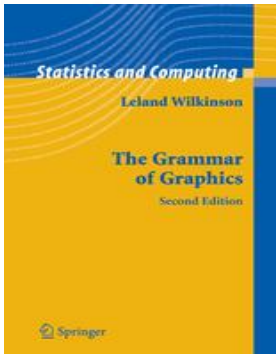


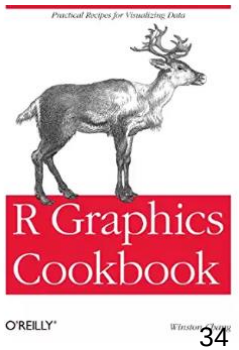












START WRITING TODAY


And you will find your
unique VOICE.

Respect & Serve Your Audience

Create and share visual stories with intensity and passion, and you can

MAKE AN IMPACT!

Alyssa Columbus. @alycolumbus. acolumbu@uci.edu