Data Analytics Workshop

Ryan Sandefer, PhD

The College of St. Scholastica

About the presenters

Ryan Sandefer, PhD

- Associate Professor and Assistant Vice President for Academic Affairs at The College of St. Scholastica in Duluth, MN
- Master's in Political Science from the University of Wyoming
- PhD in Health Informatics from the University of Minnesota
- An active member of AHIMA
  - Prior chair of the Council for Excellence in Education
  - Prior chair of the Health Information Management-Reimagined taskforce
- An active member of AMIA

Overall Goal

Communicating with Data

Agenda

Session 1

- Data visualization and telling stories through data analysis

Session 2

- Basic Data Analytics with MS Excel
  - Data Visualization and Data Summaries

Session 2

- Data Analytics with Tableau
  - Data Visualization and Dashboard Development

I have a story to tell

- In St. Louis County, MN 4% of the population is uninsured with an unemployment rate of 4.6% and 79% graduating from high school. 29% of adults are obese, 18% of adults smoke, and 9% of the population have diabetes. Average life expectancy is 78.8 years old.

What information did you gather from this story that allows you to derive knowledge for decision-making?

http://www.countyhealthrankings.org/app/minnesota/2019/rankings/st-louis/county/outcomes/overall/snapshot

My story is flawed and incomplete

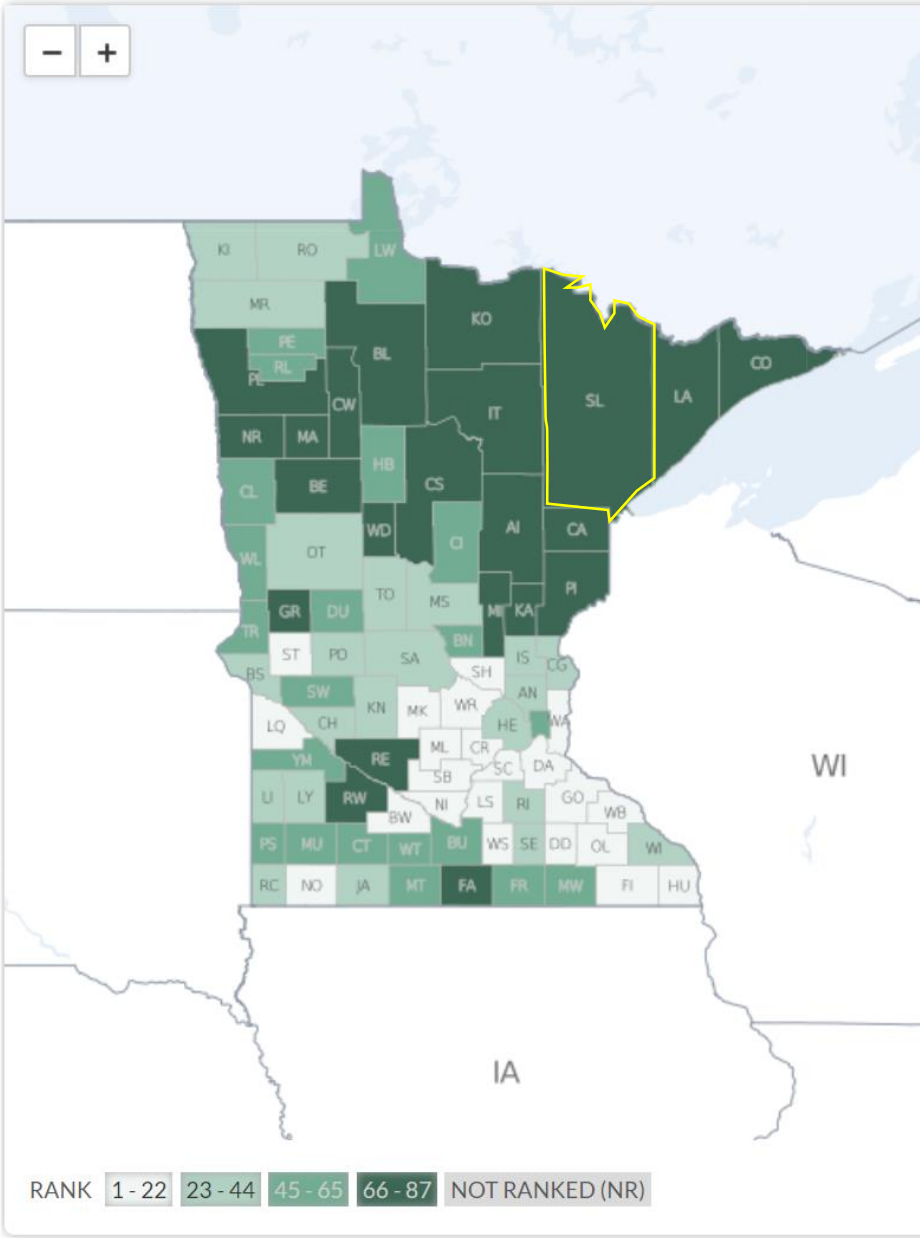
- What is the objective of my story?
- Why is this data important?
- How does St. Louis County compare to other counties?
- Is the presentation of the data effective and meaningful?

Let’s Try Again

Investigate the need for a Diabetes Education Program in St. Louis County, MN

	St Louis County	Rank in MN *Out of 87*	MN State Rates	Top US Performer
Diabetes Prevalence	9%	41st	8%	9%
Adult Obesity	29%	22nd	28%	26%
Uninsured Rate	4%	15th	5%	6%
Overall Health Outcomes		76th		

Overall Rankings in Health Outcomes



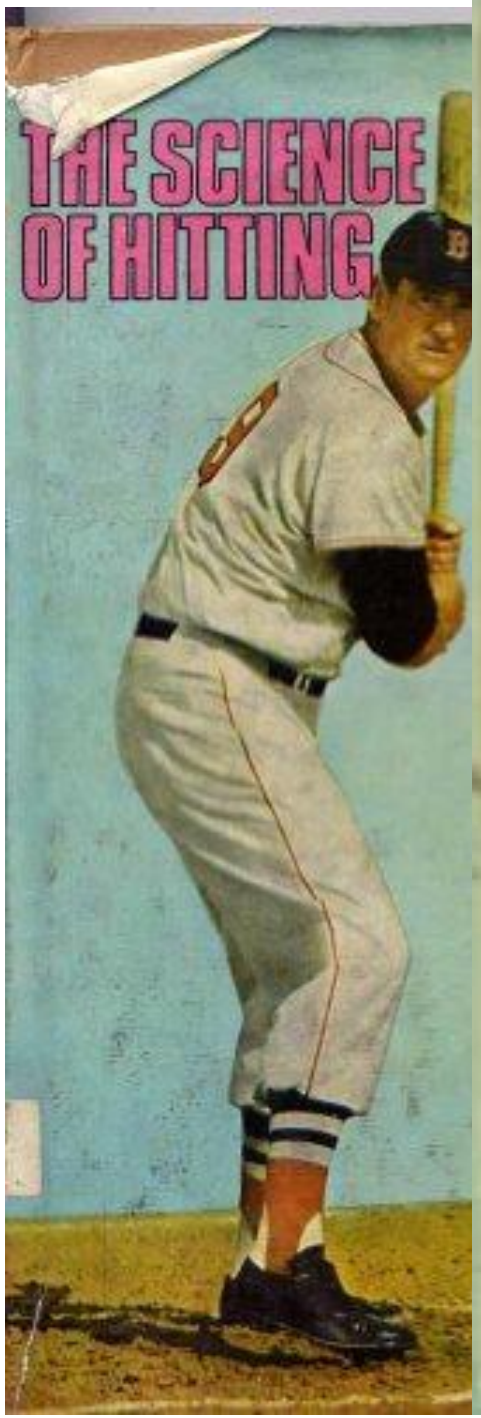
http://www.countyhealthrankings.org/app/arkansas/2019/overview

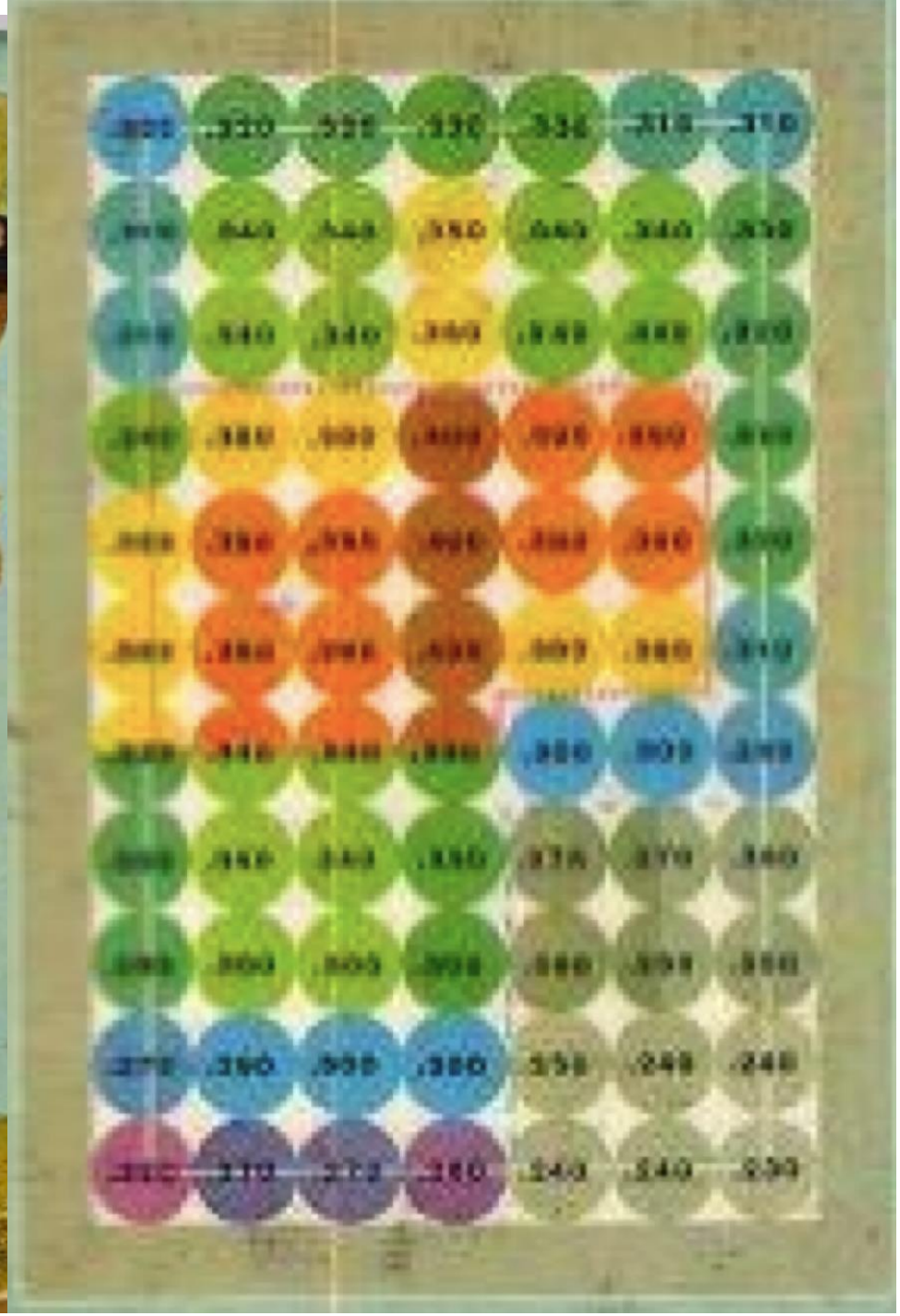
Communicate with a Story

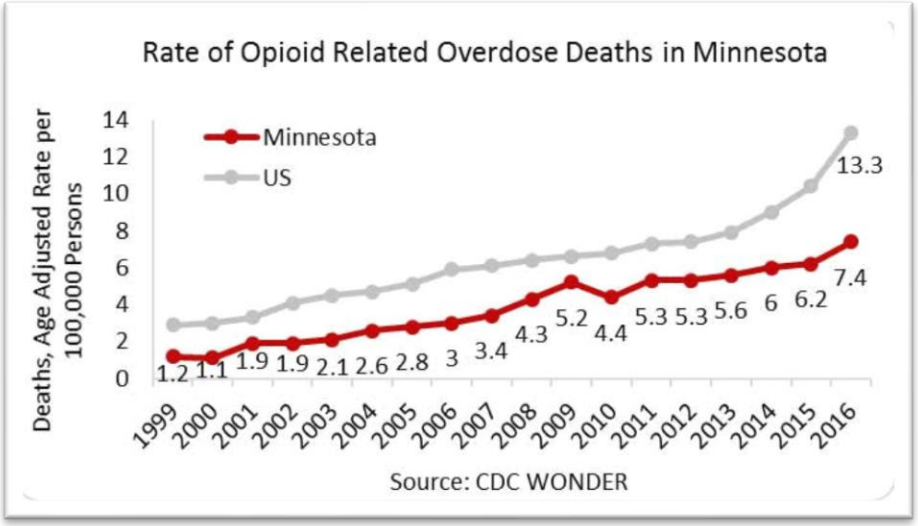
- You should strive to tell a story with your data
- Don’t just measure something for the sake of measuring something. There should be a clear purpose!
- There should be a clear start and end
- Data visualization helps communicate a story effectively
- Here’s a good example: https://youtu.be/6xsvGYIxJok

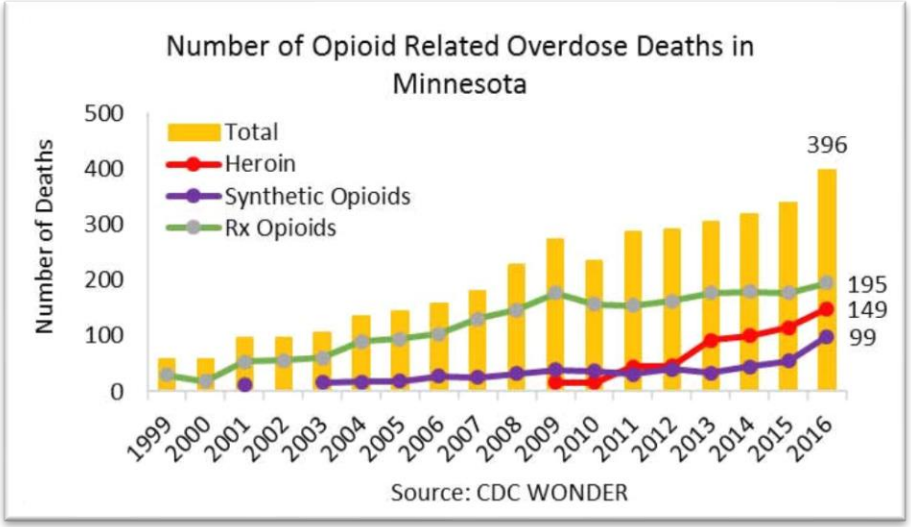
We are going to discuss a 4 steps for storytelling with data

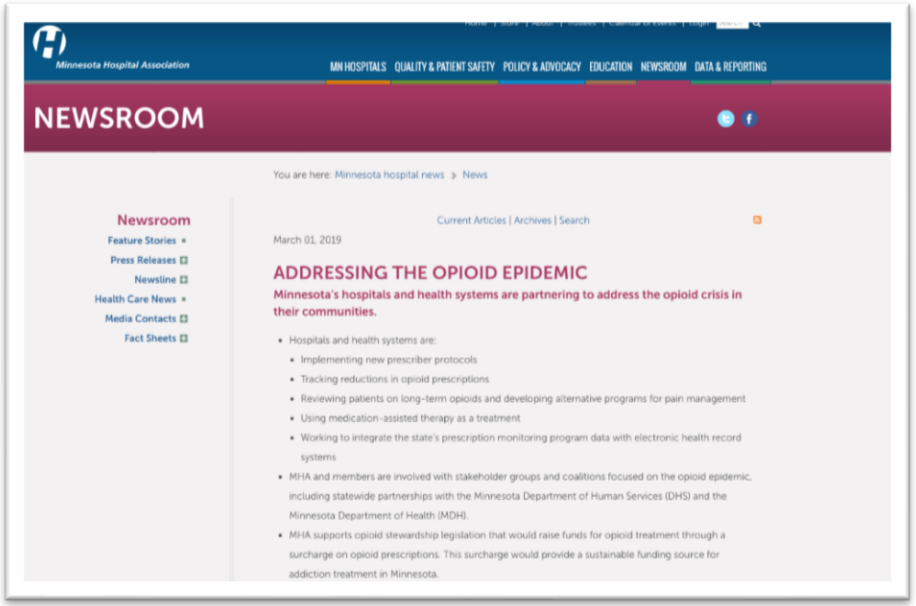
1. Pose a *good* question
2. Define *good* measures
3. Determine a *good* data source
4. Create a *meaningful* message


THE SCIENCE
OF HITTING




Rate of Opioid Related Overdose Deaths in Minnesota
Deaths, Age Adjusted Rate per 100,000 Persons
Minnesota
US
1.2 1.1 1.9 1.9 2.1 2.6 2.8 3 3.4 4.3 5.2 4.4 5.3 5.3 5.6 6 6.2 7.4
13.3
Source: CDC WONDER


Number of Opioid Related Overdose Deaths in Minnesota
Number of Deaths
Total
Heroin
Synthetic Opioids
Rx Opioids
396
195
149
99
Source: CDC WONDER


Minnesota Hospital Association
MN HOSPITALS QUALITY & PATIENT SAFETY POLICY & ADVOCACY EDUCATION NEWSROOM DATA & REPORTING
NEWSROOM
You are here: Minnesota hospital news > News
Newsroom
Feature Stories
Press Releases
Newsline
Health Care News
Media Contacts
Fact Sheets
Current Articles | Archives | Search
March 01, 2019
ADDRESSING THE OPIOID EPIDEMIC
Minnesota's hospitals and health systems are partnering to address the opioid crisis in their communities.
Hospitals and health systems are:
Implementing new prescriber protocols
Tracking reductions in opioid prescriptions
Reviewing patients on long-term opioids and developing alternative programs for pain management
Using medication-assisted therapy as a treatment
Working to integrate the state's prescription monitoring program data with electronic health record systems
MHA and members are involved with stakeholder groups and coalitions focused on the opioid epidemic, including statewide partnerships with the Minnesota Department of Human Services (DHS) and the Minnesota Department of Health (MDH).
MHA supports opioid stewardship legislation that would raise funds for opioid treatment through a surcharge on opioid prescriptions. This surcharge would provide a sustainable funding source for addiction treatment in Minnesota.


DEPARTMENT OF HUMAN SERVICES
Working Together to End the Opioid Epidemic
In order to address the opioid crisis gripping the state, the Minnesota Department of Human Services is working with state and community partners to increase awareness, decrease the number of persons who develop opioid use disorder and reduce the harms resulting from opioid addiction.
Statewide data from the Minnesota Department of Health related to opioid use, misuse, and overdose death prevention
Know the Dangers: Community-focused opioid prevention campaign
Efforts to date: Attacking the problem from all angles
As the state authority on substance abuse disorder prevention and treatment, the Department of Human Services has a big role in tackling this problem.
DHS efforts include:
A community response to opioids: grants to prevent and treat opioid use disorder
Supporting communities: State Targeted Response to the Opioid Crisis
Expanding proven treatments: Medication-Assisted Treatment

1) Create a question

- What question are you hoping to answer with your data?
  - Try to avoid complex questions
  - Keep in mind what you want to measure and compare and try to capture this in your question

Bad:

Are hospitals impacted by patients diagnosed with mental health disorders over time?

Good:

Is the percentage of patients admitted to the ED with mental health disorders different across the past 6 months?

Create a question

Is there a difference in average years of life lost in St. Louis County, MN compared to other counties in MN and compared to the overall state rate?

2) Define what you want to measure

- Dependent variable
  - The thing being measured
  - E.g., Total cost of transports, # of ED patients with MH disorder
- Independent variable
  - The thing being compared
  - E.g., Months, pre-post treatment

The dependent variable can be compared across each level of the independent variable

Define what you want to measure

- E.g., Decrease # of Emergency Department Visits with a Behavioral Health Diagnosis in next 6 months

- **DV**: # of ED Visits
- **IV**: Months

- Considerations:
  - Define an ED visit
  - Define a behavioral health diagnosis
  - Is the count an appropriate metric? Should it be a proportion instead?

Define what you want to measure

- For proportions, define the following:

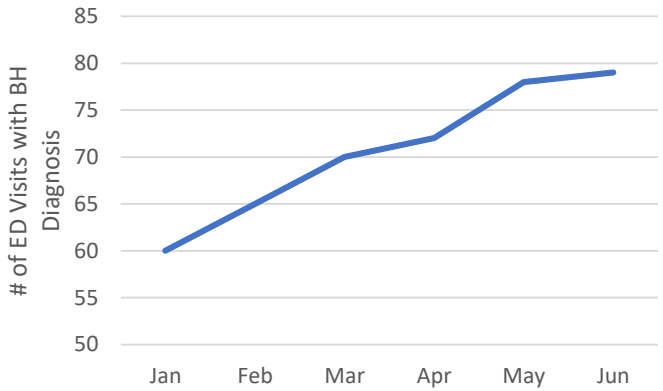
$$\frac{\textbf{Numerator}}{\textbf{Denominator}}$$

- Numerator= top number of a fraction
  - Total # of Patients with Opioid Addiction in Duluth
- Denominator= bottom number of a fraction
  - Total # of Patients in Duluth

Define what you want to measure

What story do you want to tell?

	ED Visits with BH	Total ED Visits	Proportion
Jan	60	150	0.400
Feb	65	165	0.394
Mar	70	172	0.407
Apr	72	175	0.411
May	78	193	0.404
Jun	79	199	0.397



Determine what you're going to do with the measure

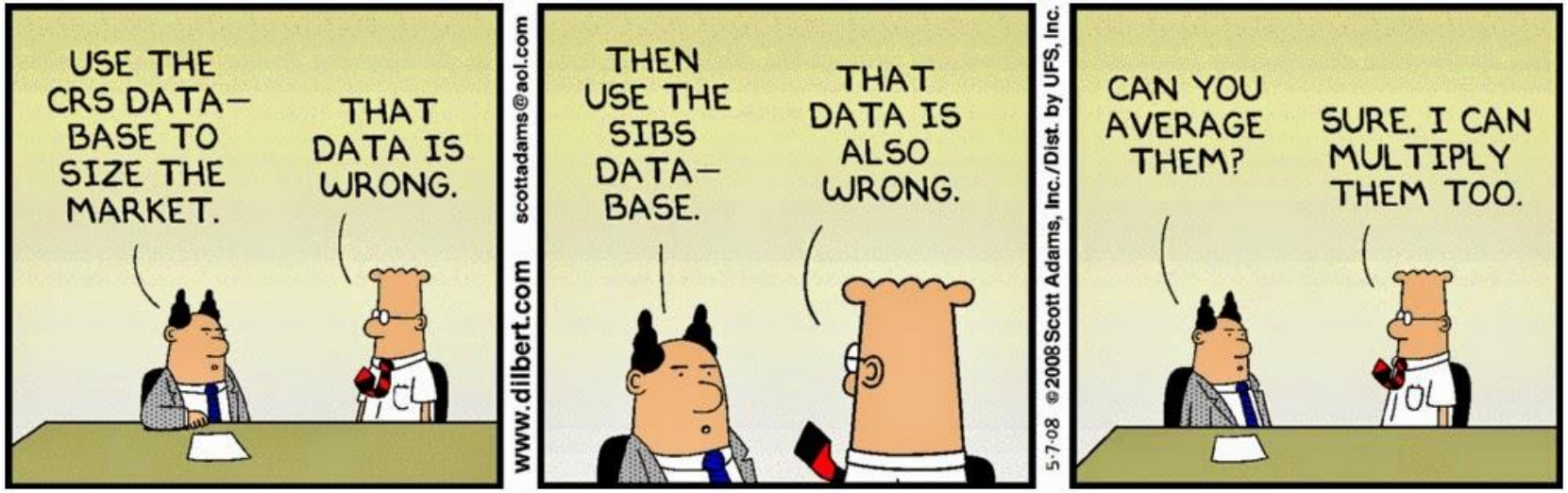
- Examine differences:
  - Over time
  - Pre and post intervention
  - Between groups (e.g., location A vs. location B)

- How will the differences be compared?
  - Average
  - Median
  - Percentage
  - Counts

- E.g., Decreased PHQ-9 Scores upon mental health follow-up
  - Compare average difference in PHQ-9 pre and post mental health treatment

Average isn't always to best way to describe the data!

- Such because you can, doesn't mean you should
- If it looks like a number, doesn't mean it is a number
  - Male = 1
  - Female = 2
- The average can be misleading if the data is skewed



Define what you want to measure

- There’s no need to reinvent the wheel. Often times, data or metrics are available and can be repurposed.
- Other times, you need to collect your own data and develop your own metrics.
- Knowing where the data resides, is a good a start!

- We will talk about both options…

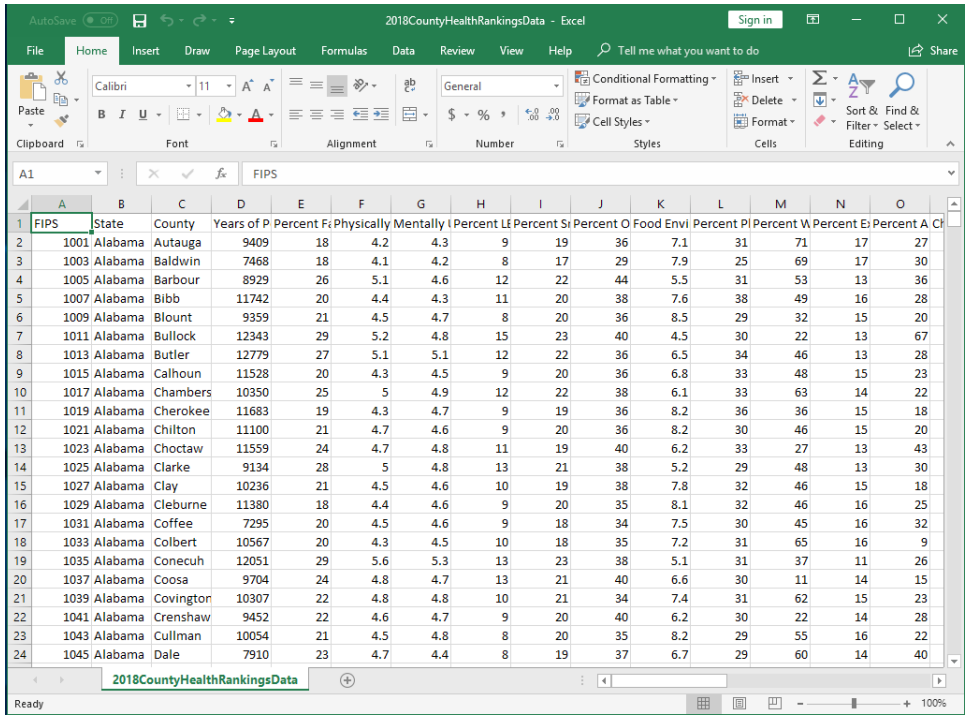
3) Where is the data?

- Healthcare is complex and the data is complex

1. Determine if the data you want is from an internal or external source
2. Work closely with your IT department or community partners to provide you with data
   - Say what you want
   - When you get what you want, don’t assume it is correct
   - Be critical of your data

The data we will be using

- The data was acquired from the 2019 County Health Rankings: http://www.countyhealthrankings.org/



FIPS	State	County	Years of P	Percent Fa	Physically	Mentally I	Percent LE	Percent Sı	Percent O	Food Envi	Percent Pl	Percent W	Percent Ea	Percent A	Ch
1001	Alabama	Autauga	9409	18	4.2	4.3	9	19	36	7.1	31	71	17	27	
1003	Alabama	Baldwin	7468	18	4.1	4.2	8	17	29	7.9	25	69	17	30	
1005	Alabama	Barbour	8929	26	5.1	4.6	12	22	44	5.5	31	53	13	36	
1007	Alabama	Bibb	11742	20	4.4	4.3	11	20	38	7.6	38	49	16	28	
1009	Alabama	Blount	9359	21	4.5	4.7	8	20	36	8.5	29	32	15	20	
1011	Alabama	Bullock	12343	29	5.2	4.8	15	23	40	4.5	30	22	13	67	
1013	Alabama	Butler	12779	27	5.1	5.1	12	22	36	6.5	34	46	13	28	
1015	Alabama	Calhoun	11528	20	4.3	4.5	9	20	36	6.8	33	48	15	23	
1017	Alabama	Chambers	10350	25	5	4.9	12	22	38	6.1	33	63	14	22	
1019	Alabama	Cherokee	11683	19	4.3	4.7	9	19	36	8.2	36	36	15	18	
1021	Alabama	Chilton	11100	21	4.7	4.6	9	20	36	8.2	30	46	15	20	
1023	Alabama	Choctaw	11559	24	4.7	4.8	11	19	40	6.2	33	27	13	43	
1025	Alabama	Clarke	9134	28	5	4.8	13	21	38	5.2	29	48	13	30	
1027	Alabama	Clay	10236	21	4.5	4.6	10	19	38	7.8	32	46	15	18	
1029	Alabama	Cleburne	11380	18	4.4	4.6	9	20	35	8.1	32	46	16	25	
1031	Alabama	Coffee	7295	20	4.5	4.6	9	18	34	7.5	30	45	16	32	
1033	Alabama	Colbert	10567	20	4.3	4.5	10	18	35	7.2	31	65	16	9	
1035	Alabama	Conecuh	12051	29	5.6	5.3	13	23	38	5.1	31	37	11	26	
1037	Alabama	Coosa	9704	24	4.8	4.7	13	21	40	6.6	30	11	14	15	
1039	Alabama	Covington	10307	22	4.8	4.8	10	21	34	7.4	31	62	15	23	
1041	Alabama	Crenshaw	9452	22	4.6	4.7	9	20	40	6.2	30	22	14	28	
1043	Alabama	Cullman	10054	21	4.5	4.8	8	20	35	8.2	29	55	16	22	
1045	Alabama	Dale	7910	23	4.7	4.4	8	19	37	6.7	29	60	14	40	

About the data

- Reference the data dictionary that was shared

County Health Rankings & Roadmaps
Building a Culture of Health, County by County

A Robert Wood Johnson Foundation program

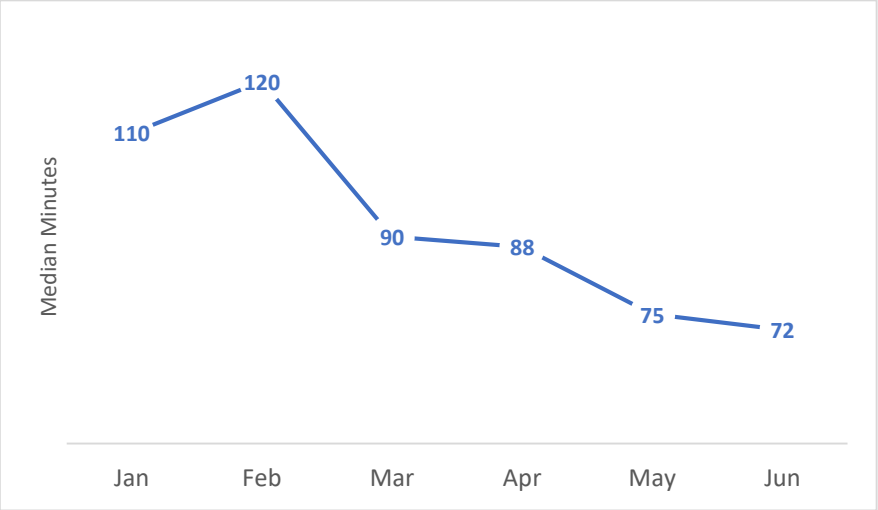
A collaboration between the Robert Wood Johnson Foundation and the University of Wisconsin Population Health Institute

This Excel file contains the rankings and data details for the measures used in the 2018 *County Health Rankings*.
1) Outcomes and Factors Rankings--Ranks are all calculated and reported WITHIN states
2) Outcomes and Factors SubRankings--Ranks are all calculated and reported WITHIN states
3) Ranked Measure Data--The measures themselves are listed in bold.
4) Ranked Measure Sources & Years
5) Additional Measure Data--These are supplemental measures reported on the *Rankings* web site but not used in calculating the rankings.
6) Additional Measure Sources & Years

Measure	Data Elements	Description
Geographic identifiers	**FIPS**	Federal Information Processing Standard
	State	
	County	
Premature death	**Years of Potential Life Lost Rate**	Age-adjusted YPLL rate per 100,000
	95% CI - Low	95% confidence interval reported by National Center for Health Statistics
	95% CI - High	
	Quartile	Within-state rank: 1 = top quartile, 2=second quartile, 3= third quartile, 4=bottom quartile
	Years of Potential Life Lost Rate (Black)	Age-adjusted YPLL rate per 100,000 for non-Hispanic Blacks
	Years of Potential Life Lost Rate (Hispanic)	Age-adjusted YPLL rate per 100,000 for Hispanics
	Years of Potential Life Lost Rate (White)	Age-adjusted YPLL rate per 100,000 for non-Hispanic Whites
Poor or fair health	**% Fair/Poor**	Percentage of adults that report fair or poor health
	95% CI - Low	95% confidence interval reported by BRFSS
	95% CI - High	
	Quartile	Within-state rank: 1 = top quartile, 2=second quartile, 3= third quartile, 4=bottom quartile
Poor physical health days	**Physically Unhealthy Days**	Average number of reported physically unhealthy days per month
	95% CI - Low	95% confidence interval reported by BRFSS
	95% CI - High	
	Quartile	Within-state rank: 1 = top quartile, 2=second quartile, 3= third quartile, 4=bottom quartile
Poor mental health days	**Mentally Unhealthy Days**	Average number of reported mentally unhealthy days per month
	95% CI - Low	95% confidence interval reported by BRFSS
	95% CI - High	
	Quartile	Within-state rank: 1 = top quartile, 2=second quartile, 3= third quartile, 4=bottom quartile
Low birthweight	Unreliable	Value reported but considered unreliable since based on counts of twenty or less.
	% LBW	Percentage of births with low birth weight (<2500g)
	95% CI - Low	95% confidence interval
	95% CI - High	
	Quartile	Within-state rank: 1 = top quartile, 2=second quartile, 3= third quartile, 4=bottom quartile
	% LBW (Black)	Percentage of births with low birth weight (<2500g) for non-Hispanic Blacks
	% LBW (Hispanic)	Percentage of births with low birth weight (<2500g) for Hispanics
	% LBW (White)	Percentage of births with low birth weight (<2500g) for non-Hispanic Whites
Adult smoking	**% Smokers**	Percentage of adults that reported currently smoking
	95% CI - Low	95% confidence interval reported by BRFSS
	95% CI - High	

4) Translate data into meaningful information

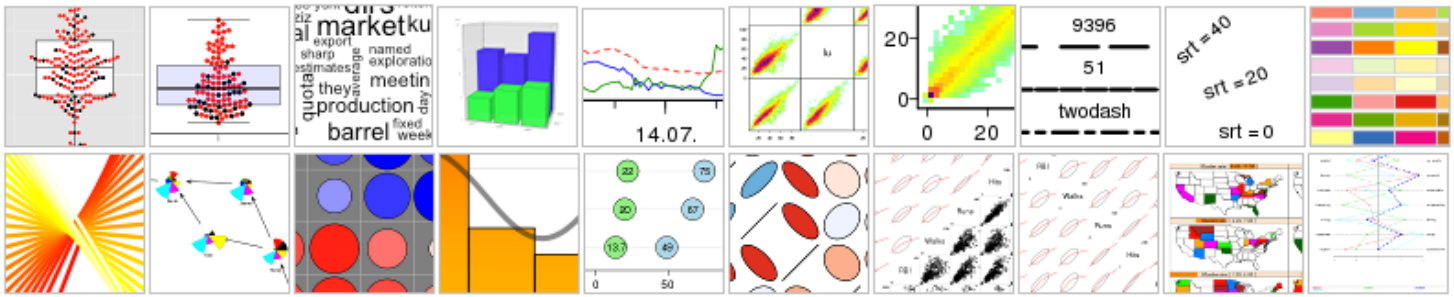
- Know your purpose and audience
- Use the space wisely!
  - Most readers read the top left of a screen first, so make the important content span that part of the screen
- Make sure you understand what type of device the viewer will be using
  - This will impact the size of your dashboard
- Don't overcrowd the display
- Add interactivity to encourage exploration
- E.g., Median time spent in the ER prior to transfer to inpatient setting in past six months

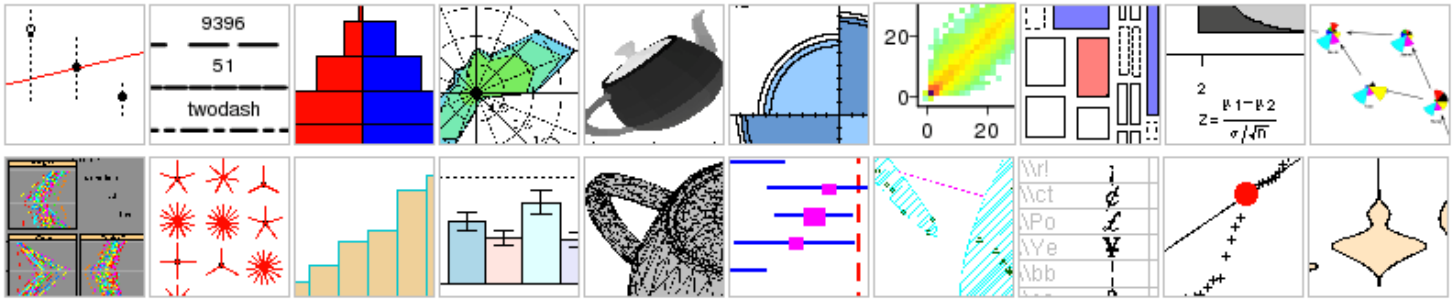


https://onlinehelp.tableau.com/current/pro/desktop/en-us/dashboards_best_practices.htm

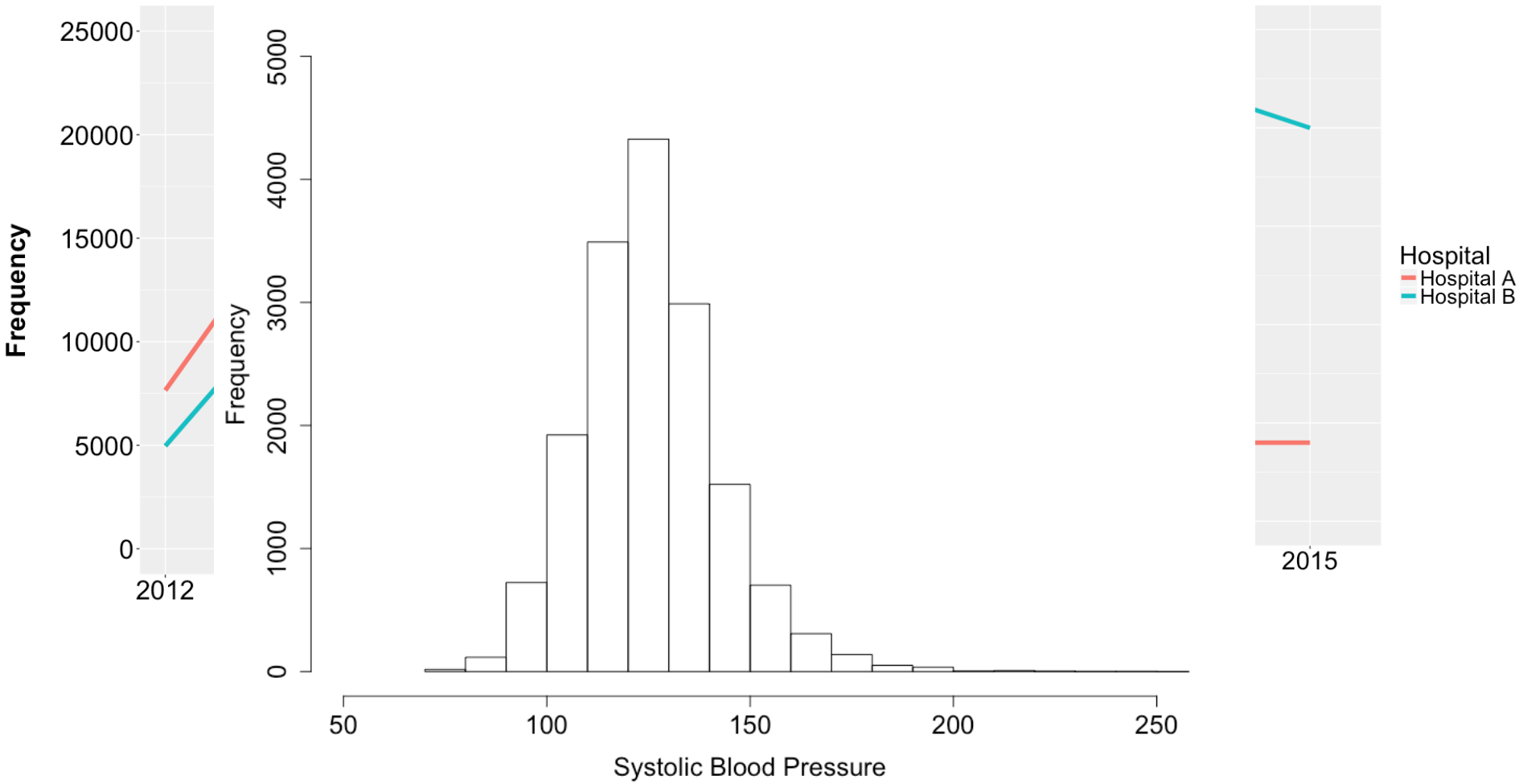
Data Visualization

- Support the transition of data into information through a visual context
- Graphs are a paramount component

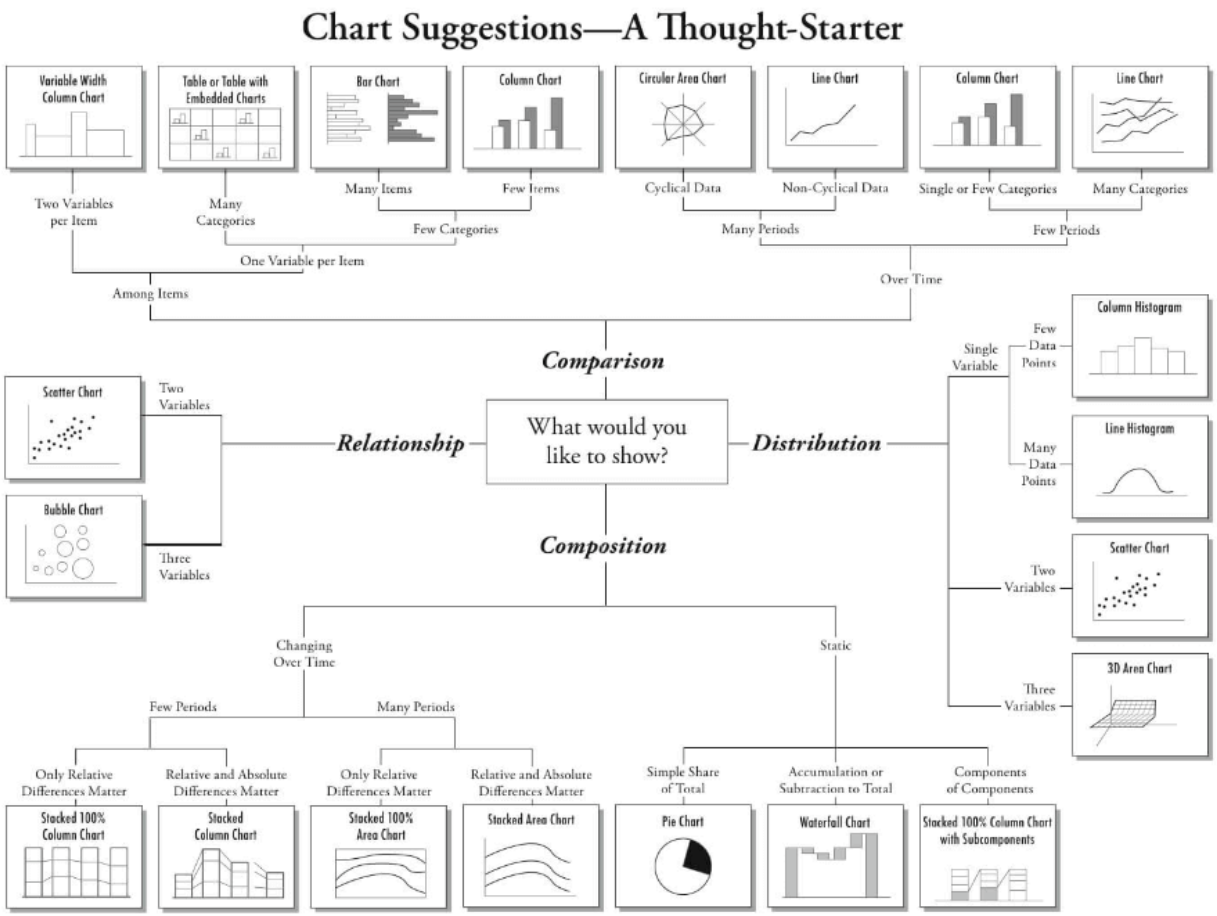




Choosing a graph!



Choosing a graph!



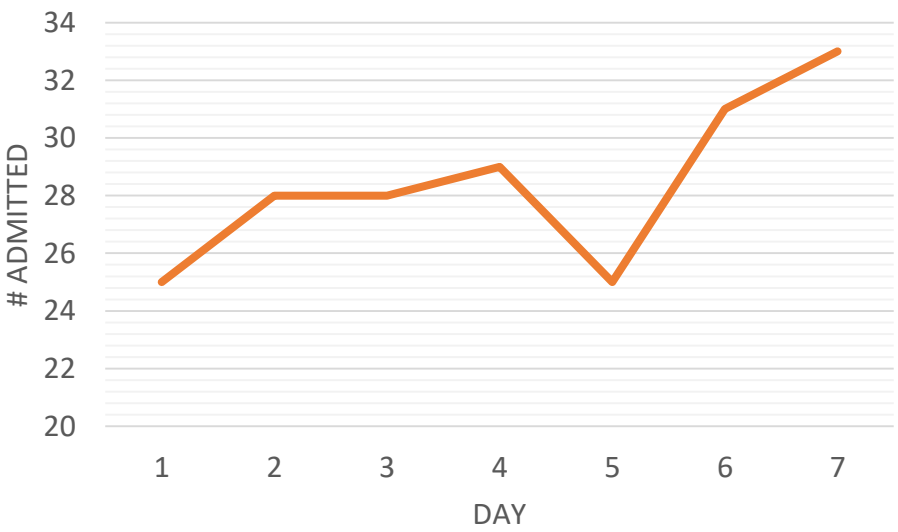
https://i1.wp.com/www.tatvic.com/blog/wp-content/uploads/2016/12/Pic_2.png

Is a picture always preferred?

- What is the general trend of the # admitted over time?
- What # was admitted on day 4?
- What day had the lowest # admitted?

Day	# Admitted
1	25
2	28
3	28
4	29
5	25
6	31
7	33

VS.



Don’t ignore your intent!

- If you create a visualization that has nothing to do with your original intent, it won’t be very meaningful
- Always ask yourself, “Why is this important and how does it relate back to what I’m doing?”
- E.g., If your intent is to improve provider awareness to improve referrals to mental health providers for care coordination, would you need to know the current number of referrals? Would you need to know incarceration rate?

Introduction to Tableau

- Tableau Software is an American computer software company headquartered in Seattle, WA, USA. It produces a family of interactive data visualization products focused on business intelligence.

- The products can query relational databases, cubes, cloud database, and spreadsheets and then generates a number of graph types that can be combined into dashboards and shared over a computer network or the internet

Tableau Products

- Tableau Desktop
  - An application that allows you to drag and drop fields to analyze data. Users can visualize data and create dashboards.
- Tableau Server
  - Offers browser-based analytics that anyone can use for business intelligence.
- Tableau Online
  - A hosted version of tableau server that allows users to share analytics and insights with anyone.
- Tableau Public
  - A free version that allows users to use the basic functions of Tableau Desktop (This is what we will use!)

A note about Tableau Public

- The public version of Tableau allows you to publish a dashboard online.
- Do NOT upload and use data with PHI into Tableau Public as you may unintentionally share this data with the public
- Tableau Desktop and Server are proprietary tools that can be used when working with data with PHI

Why use Tableau?

- It’s easy to use
  - Does not require the use of a scripting language
- Can work with many data sources
  - Excel, text files, databases, data marts, etc.
  - Can connect directly to a data source
- It’s fast!
  - Can handle very large datasets efficiently
- Capabilities
  - Create interactive displays
  - Easy to interpret graphs

Weaknesses of Tableau

- Not very comprehensive
  - Doesn’t have data mining algorithms built in
  - Can not do predictive analytics
  - Can’t be integrated with other applications because the software is proprietary
  - Does not include some visualization methods that are helpful
    - E.g., boxplots, network graphs, tree-maps, heatmaps, 3-d scatterplots

Read the Whitepaper

http://www.tableau.com/sites/default/files/media/enablinghealthcareanalyticsforbetterpatientoutcomes_eng.pdf

Tableau Resources

- Video tutorials:
  - http://www.tableau.com/learn/training
- Quick start guide:
  - http://onlinehelp.tableau.com/current/pro/desktop/en-us/help.html#quickstarts.html
- Whitepapers:
  - http://www.tableau.com/learn/whitepapers

Let’s create a story from data!

We are going to create an interactive dashboard in MS Excel to compare rates of opioid addiction, diabetes, heart disease in Duluth and other cities in MN

Conclusions

- Asking a good question is critical!
- There are sources of both internal and external data
- Tell a story with your data through visualizations!

Questions?

Thank you!

Ryan Sandefer

rsandefe@css.edu