



# Welcome to the Data Visualization Webinar

We will begin here shortly

Brought to you by the ASA, CSSA, and SSSA  
Graduate Student Committee



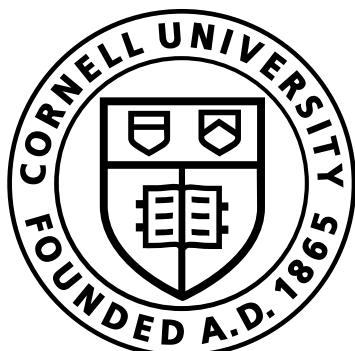
# Welcome to the Data Visualization Webinar

You should now hear introductory remark audio.

If not: 1) check your device volume 2) check your headset plug in or 3) leave webinar and reopen link

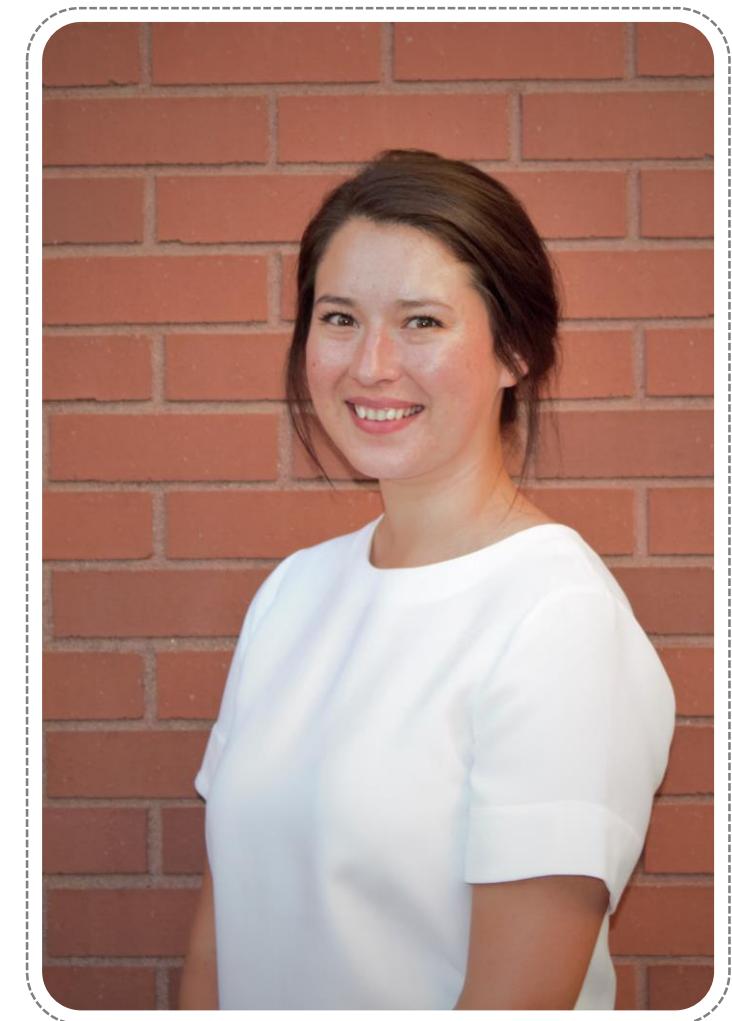
Brought to you by the ASA, CSSA, and SSSA  
Graduate Student Committee

# Dr. Shantel A. Martinez



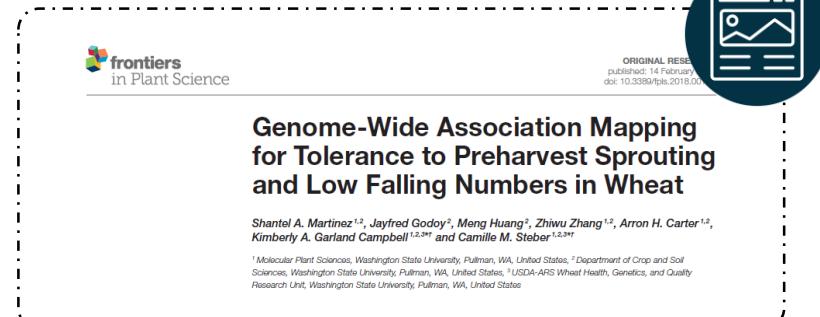
BS Bioengineering  
MS Crop Sci  
PhD Mol Plant Sci

USDA NIFA EWD  
Postdoc Fellowship



# My “Every Day” Data Viz:

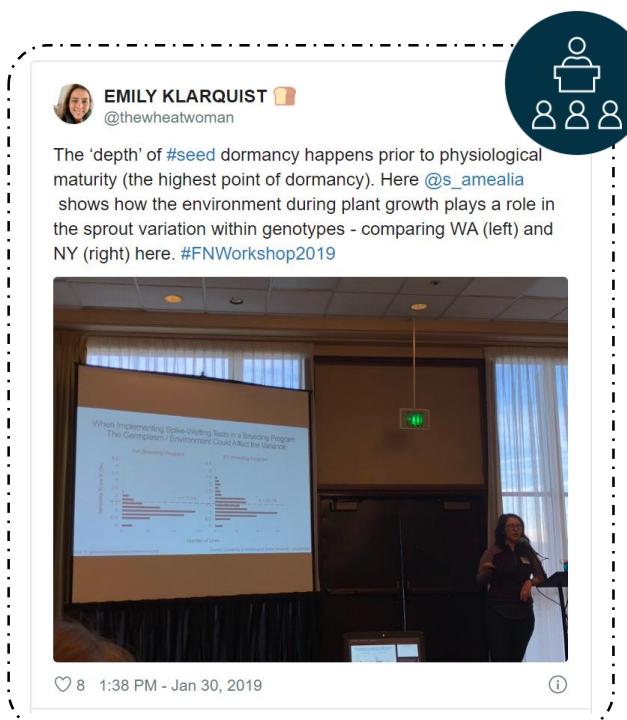
# Peer-reviewed Publications



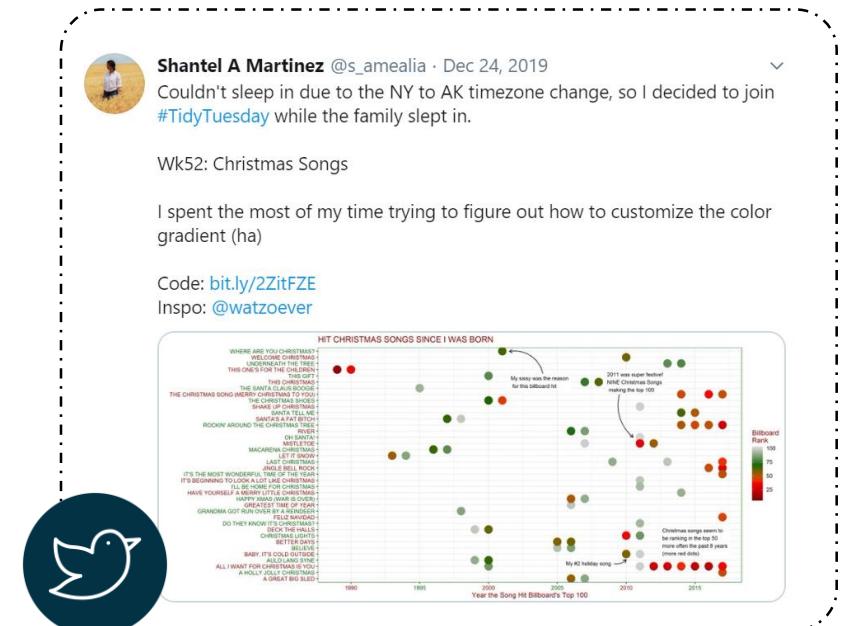
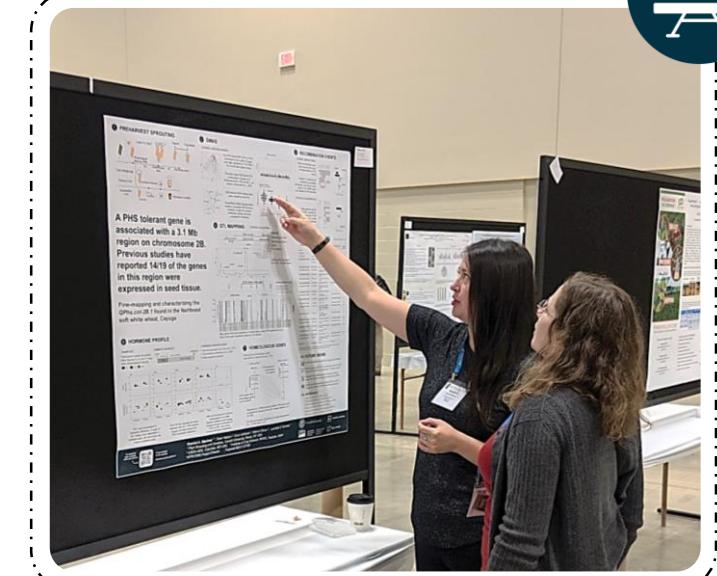
## Research Seminars



# Public Engagement @s amelia



# Academic Posters



# Goals for today's webinar:

- »»» Provide some common concepts of data visualization
- Build upon your current foundation to improve your figures
- »»» Broadly determine the type of message you want to convey
- »»» Provide resources to improve your data visualizations

# Goals for today's webinar:

## »»» Core Principles

Provide common concepts of data visualization

## Examples

Build upon your current foundation to improve your figures

## »»» Examples from One Dataset

Broadly determine the type of message you want to convey

## »»» Resources

# What is the point of visualizing data?

**Communicate complex ideas  
with clarity, precision,  
and efficiency**

# What is your comfort level with developing appropriate data visualization?

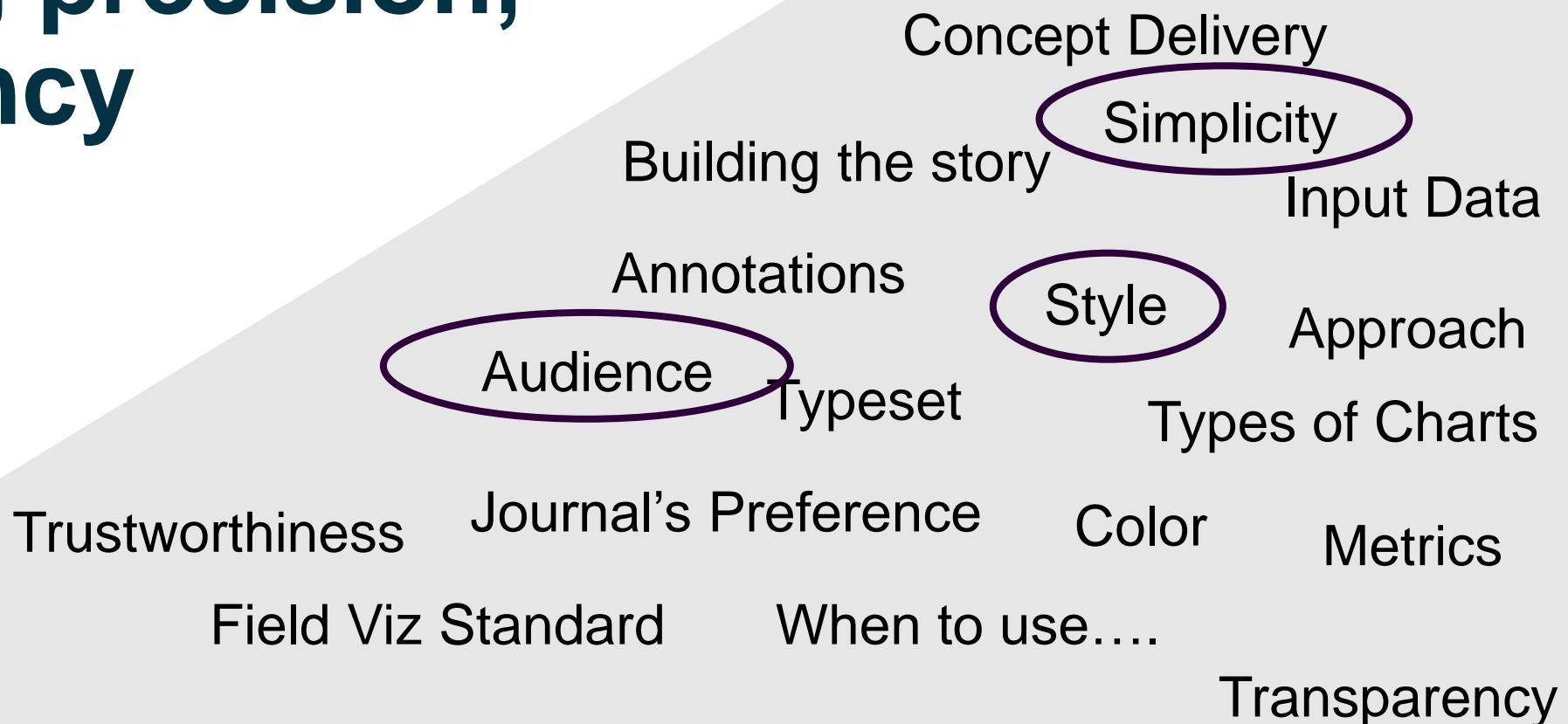
*Use the poll to answer, once it is launched*

- »»» Help!
- »»» Got the basics down
- »»» Some aspects I'm a pro
- »»» Pretty Savvy

# What is the point of visualizing data?

**Communicate complex ideas  
with clarity, precision,  
and efficiency**

**How?**



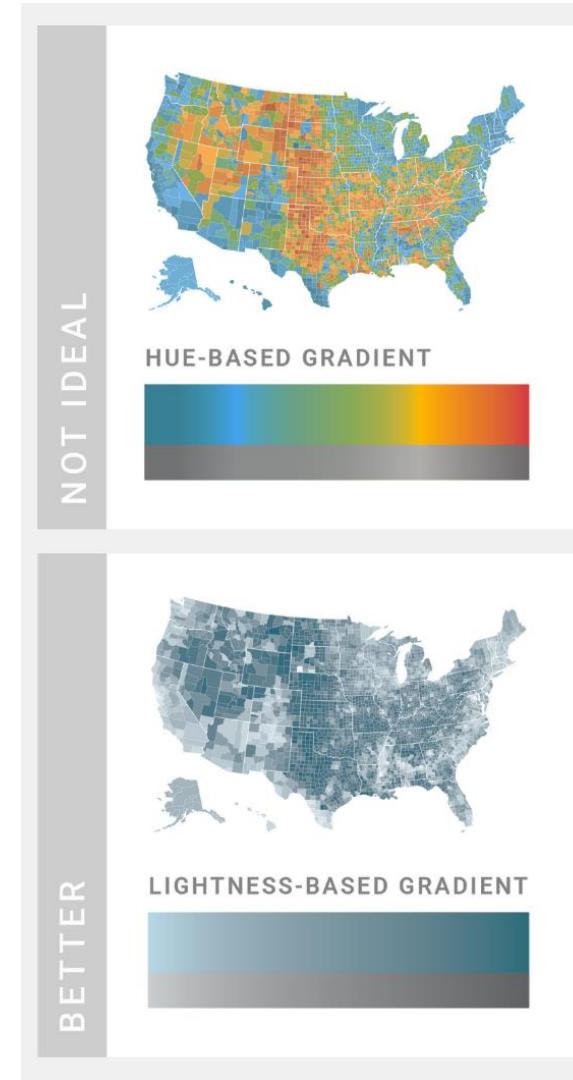
# Simplicity

Simplicity is often misunderstood to mean that we should “dumb down the message”

TIP #1

If there's a clear way to show the relationship or conclusion, then we should show it clearly

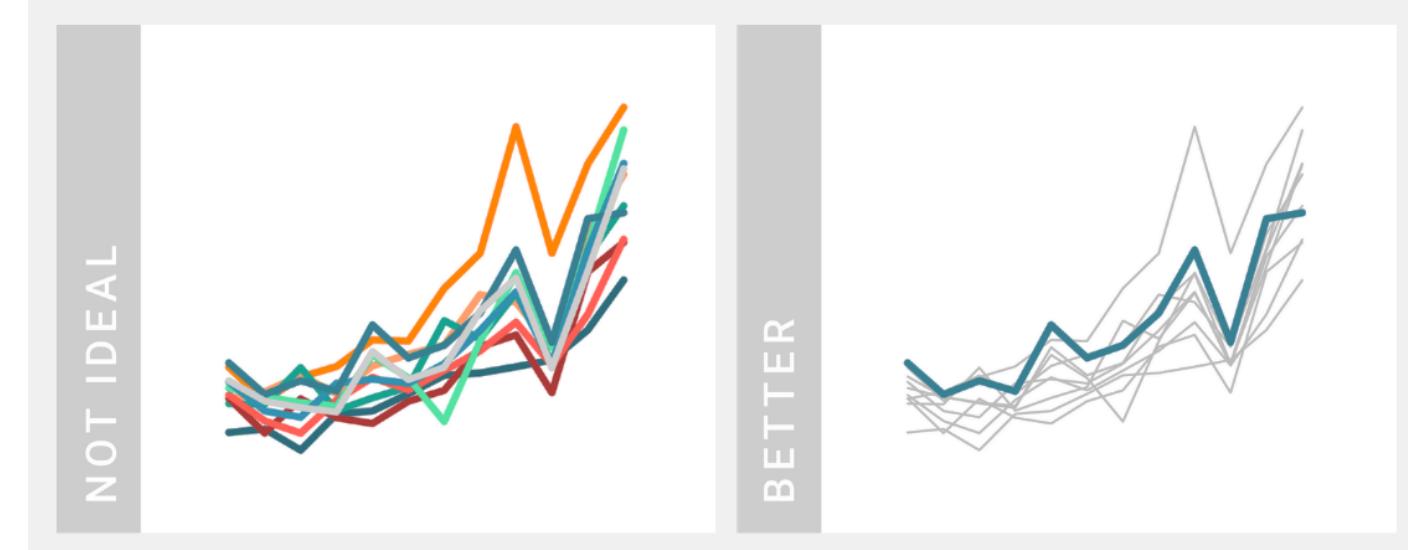
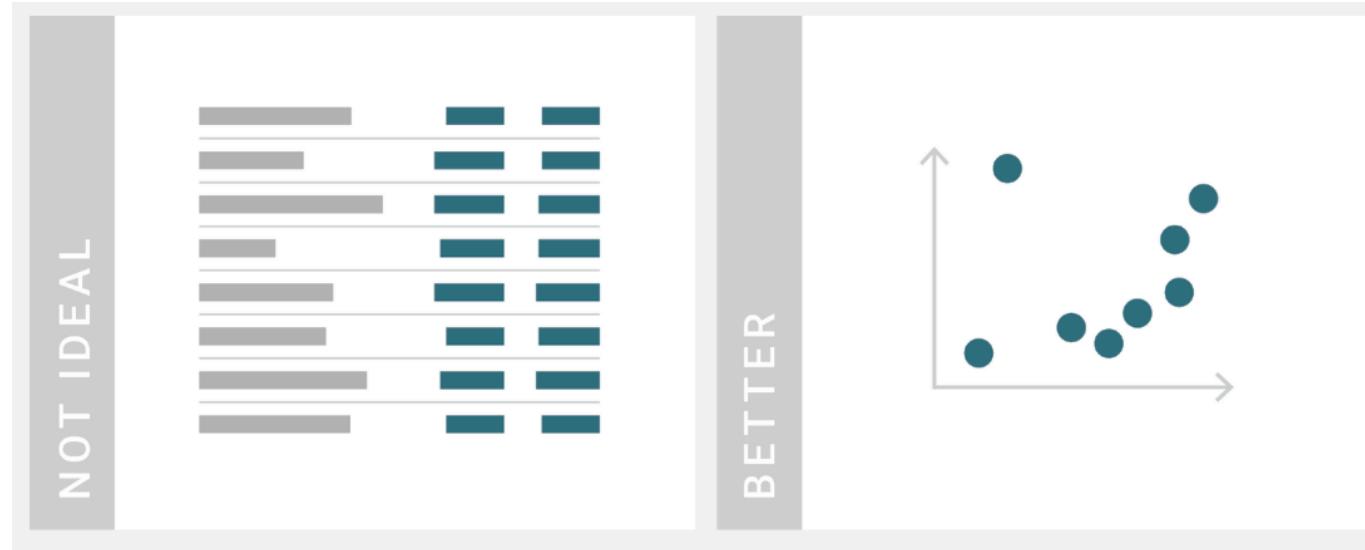
“everything should be made as simple as possible, but not simpler.” - Albert Einstein



Jones, B., (2015) DataRemixed: On Visualizing Data Well

Rost, L.C., (2018) Chartable: What to consider when choosing colors for data visualization

# Simplicity



If you needed to make a point about one line at a time

If the trend was the main point

Rost, L.C., (2019) Chartable: What to consider when creating tables

Rost, L.C., (2018) Chartable: What to consider when choosing colors for data visualization

# Style

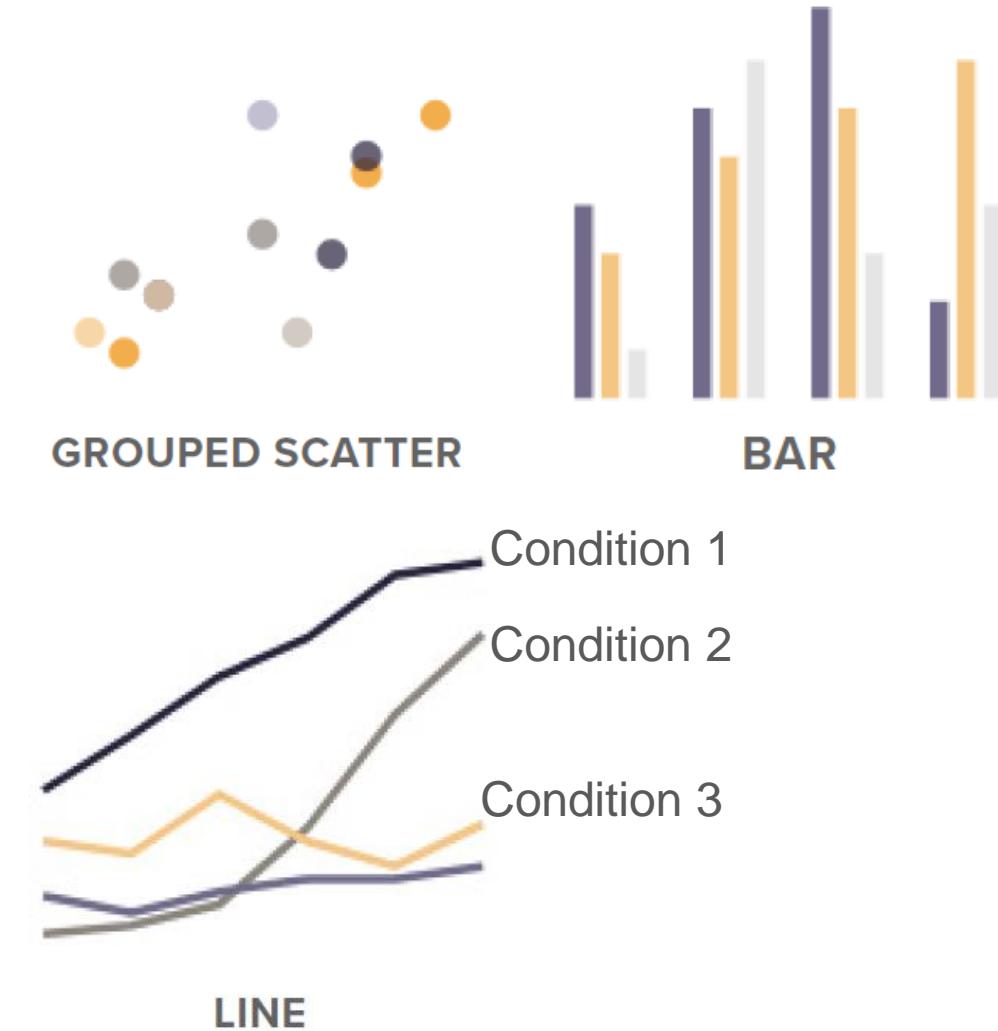
---

Clarity and beauty are not mutually exclusive

TIP #2

Focus on element consistencies to help the viewer connect topics

Color Palette & Tone



# Style

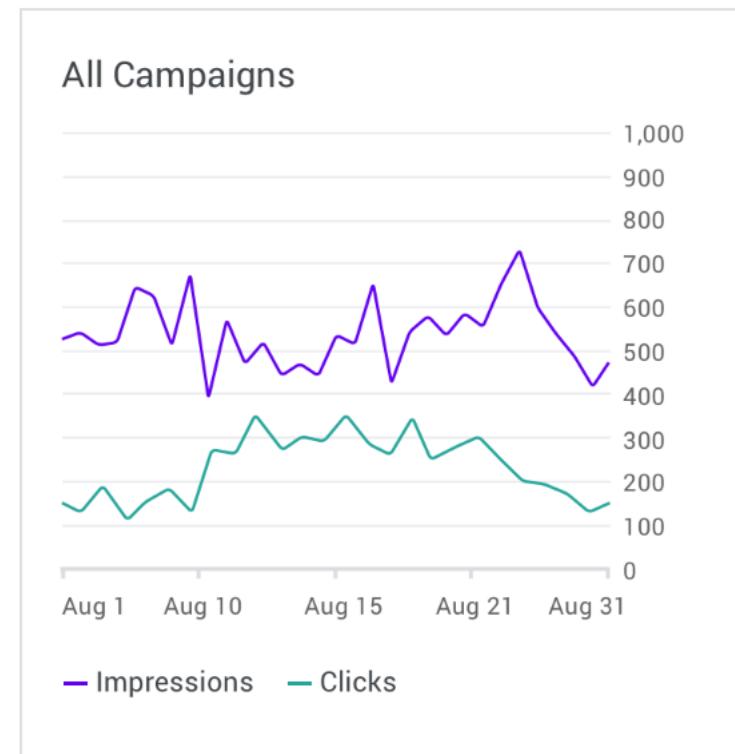
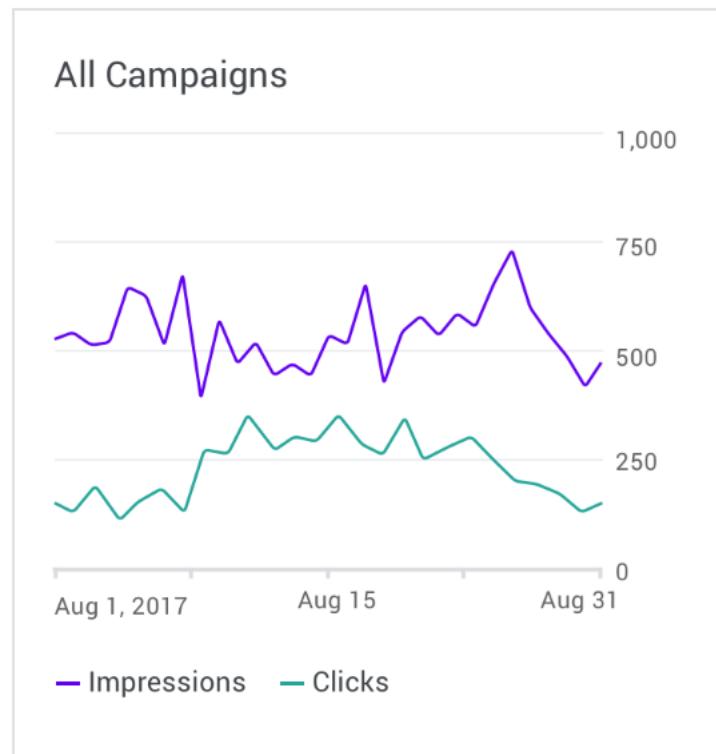
Clarity and beauty are not mutually exclusive

TIP #2

Focus on element consistencies to help the viewer connect topics

Color Palette & Tone

Line Width and Texture



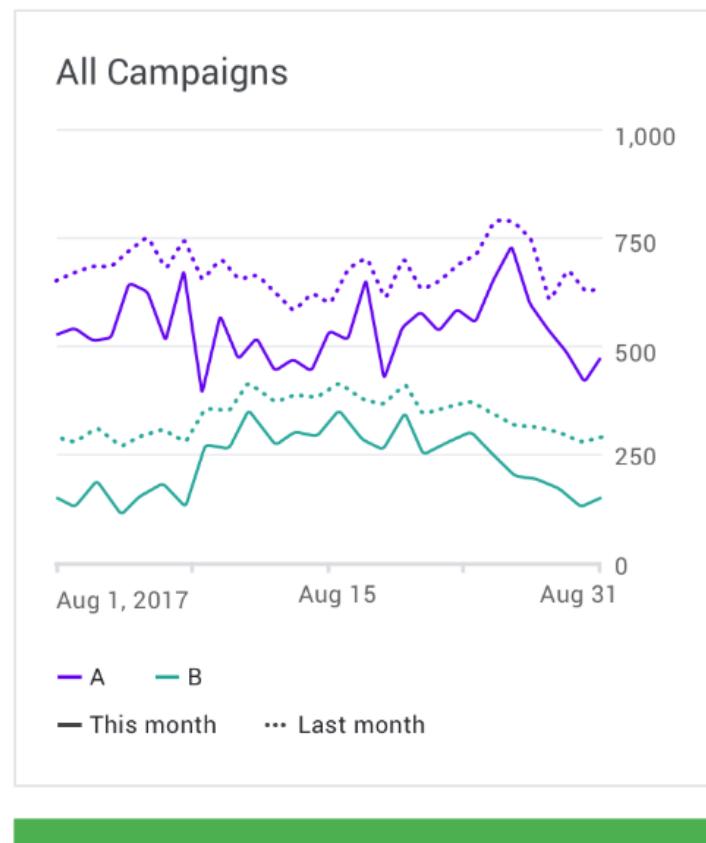
# Style

Clarity and beauty are not mutually exclusive

TIP #2  
Focus on element consistencies to help the viewer connect topics

Color Palette & Tone

Line Width and Texture



**Do.**  
Vary a line's texture to represent different data types.



**Don't.**  
Don't use different colors to show periodical variation for the same data category.

# Style

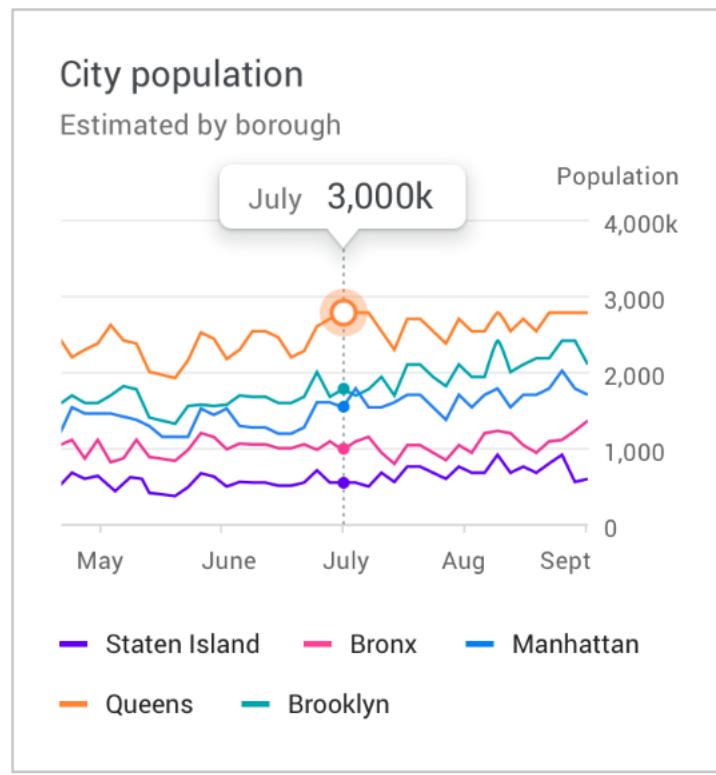
Clarity and beauty are not mutually exclusive

TIP #2  
Focus on element consistencies to help the viewer connect topics

Color Palette & Tone

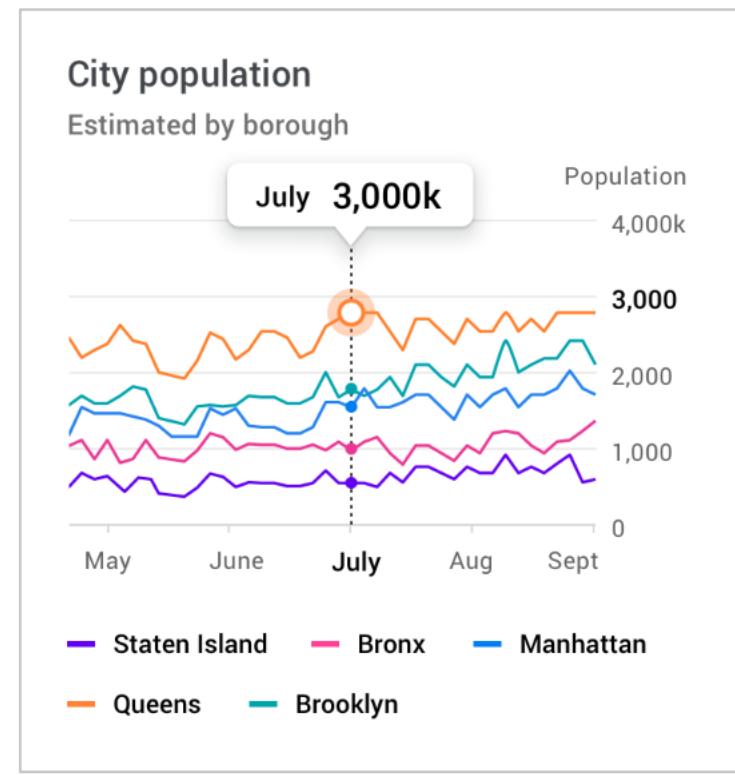
Line Width and Texture

Text position or weight



Do.

Bold used just for one or two key elements creates a balanced design.



Don't.

Bold used on too many elements can make it harder to identify important elements.

# Audience

TIP #3

Always start with defining who you're talking to?

A good visualization takes into account your audience's ability to decipher your main message

Time

How long do you have to make your point?

The Point

Focus in on the message you are trying to convey.

Detail

What amount of detail must you give to ensure clarity?

Audience:

# The researcher (you)

This is when you, and close collaborators,  
are exploring your data



Artist: Allison Horst ([2020](#))

Time

a lot of it

The Point

discover the relationships  
or conclusions

Detail

allowed to be messy

The audience members are  
**directly within your field**

They're very familiar with the  
context

# Poster Session Attendees

The data and conclusions will be presented i) accompanied by you and ii) unaccompanied, solo.

Time

matter of minutes

The Point

varies



Detail

must be stand-alone, typically with accompanied text

- i) open discussion on research
- ii) expose techniques to others
- iii) showcase the research

# Poster Session Attendees

The data and conclusions will be presented i) accompanied by you and ii) unaccompanied, solo.

Time

matter of minutes

The Point

varies

Detail

must be stand-alone, typically with accompanied text



The audience members are **broad**, but semi-tied to your topic (which is why they stop by the poster in the first place)

They may be less familiar with the background/method that leads to a result

Audience:

# Seminar Presentation

The data and conclusions will be presented by you, typically in a short and concise timeframe

Time

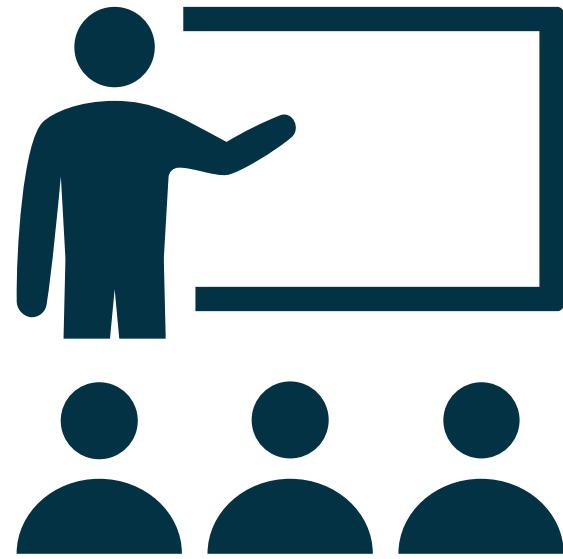
matter of seconds per graphic/table

The Point

always keep in mind the message of the graphic

Detail

focus on simplicity



The audience members are **within your field**, but does not necessarily study your project you're presenting:

They're familiar with the context, but may need a brief refresher

# Peer-reviewed Article

Communicate novel findings with often complex supporting data

Time

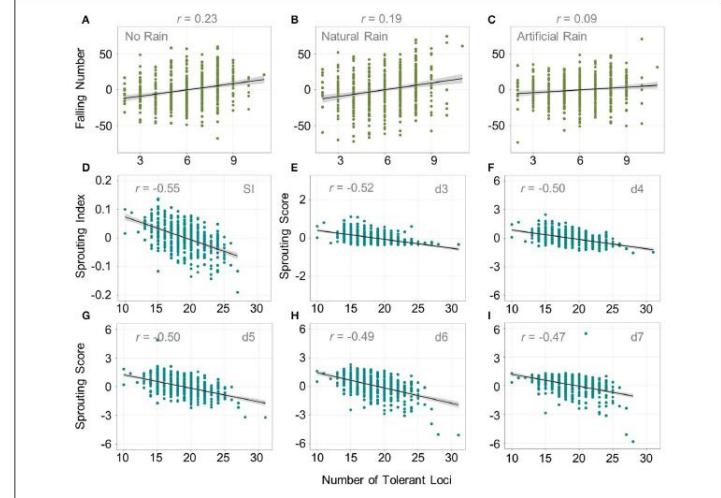
not an issue

The Point

convince the reader of the conclusions or relationships

Detail

in depth detail



**FIGURE 4 |** The effect of pyramiding multiple QFN.wsu and QPHS.wsu loci. Scatter plots of the number of favorable QFN.wsu loci versus FN BLUPs across: **(A)** in the absence of rain, **(B)** both natural rain environments combined, and **(C)** both artificial rain environments combined. Scatter plots of the number of favorable QPHS.wsu loci versus BLUPs calculated across all environments for **(D)** sprouting index, and sprouting scores on days **(E)** 3, **(F)** 4, **(G)** 5, **(H)** 6, and **(I)** 7 of miting.  $r$  is the Pearson correlation coefficient between the trait and number of tolerant loci.

## Comparative Mapping for PHS

The location of QTIN for FN and sprouting scores were compared to locations of PHS-related loci identified in 54 previous studies (Figure 5). This was done using the comparative map

The audience members can be **broad or within your field**

They may be familiar with the context, but you have the text to accompany the visualization if not

# Audience

## Public

The story is the factor that engages and makes an impression

Time

- social media – instant
- blog/news article word limit

The Point

- interest in the topic
- conclusion or relationship

Detail

broad, but ready to support with detail



# Audience

## Public

The story is the factor that engages and makes an impression

Time

- social media – instant
- blog/news article word limit

- Twitter: **limit** is 280 characters

Instagram Post: **limit** is 2,200 characters

Instagram Story: **limit** is 7 sec (photo)

broad, but ready to support with detail

The Point

Detail

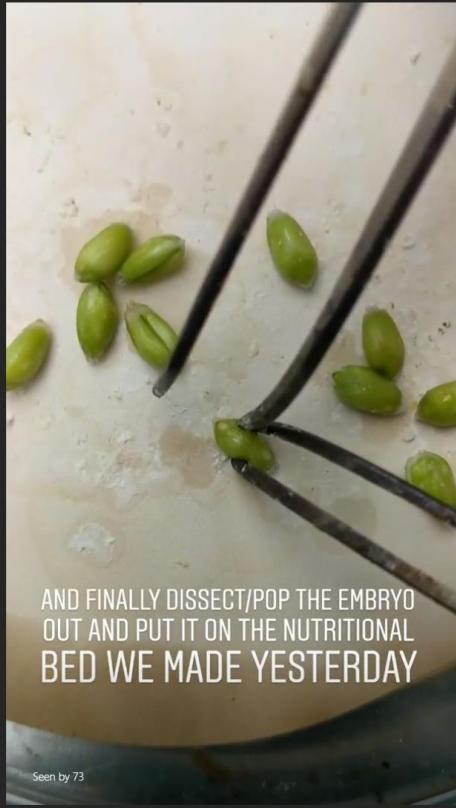


# Instagram Story:

28w



28w



27w

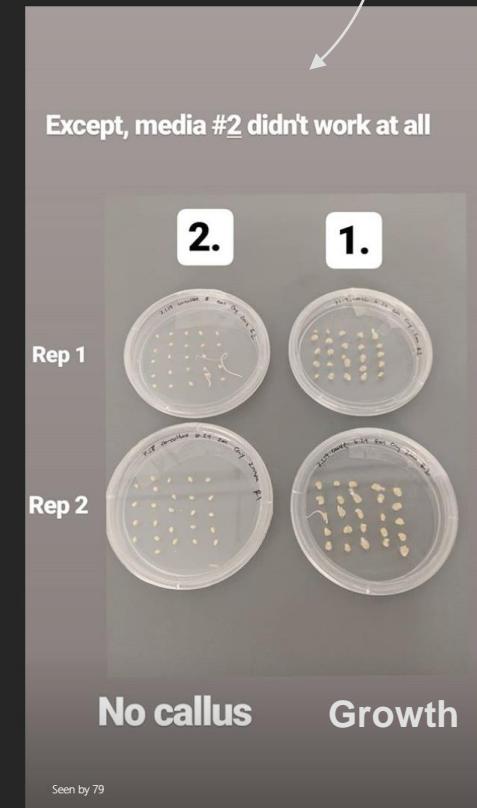


16w



Conclusion

16w



----- The Story -----

Graphic

# Audience

## Public

The story is the factor that engages and makes an impression

The audience members are likely to be **broad**

I always frame my message to my family members. People I hold in high regard, but don't study the same material I do.



# Goals for today's webinar:

»»» Core Principles

»»» Examples from Others

»»» Examples from One Dataset

Broadly determine the type of message you want to convey

»»» Resources

# To create visualizations, what tool do you use the most?

*Unfortunately, **only one answer** can be selected in the poll*

Microsoft



Python



R



SAS



Something else



# Same data: different audiences, different purpose

Data Exploration



Presentation



Poster

Public



Journal Article



NOTE: the point of these next few slides is not how to read each graph...  
the point is how they differ for each audience member

# Data Exploration

The Point

Explore different ways of viewing the data to discover the relationships or conclusions

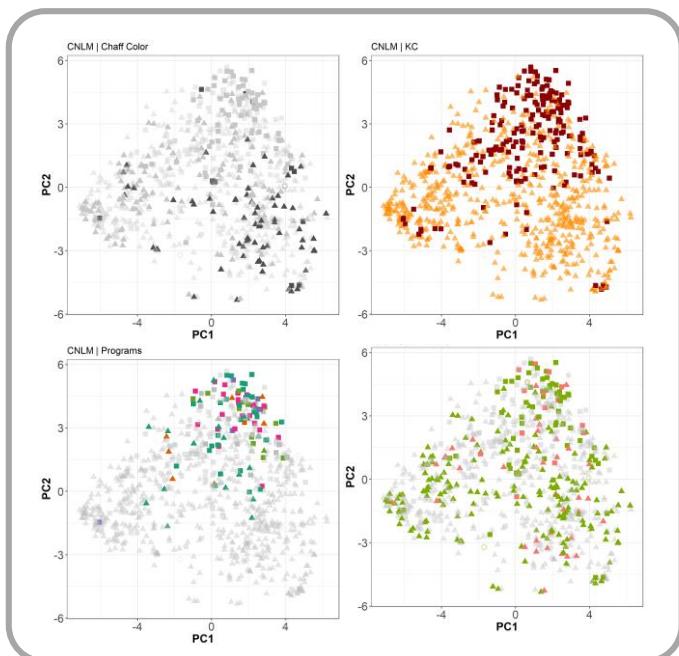
Time

a lot of it

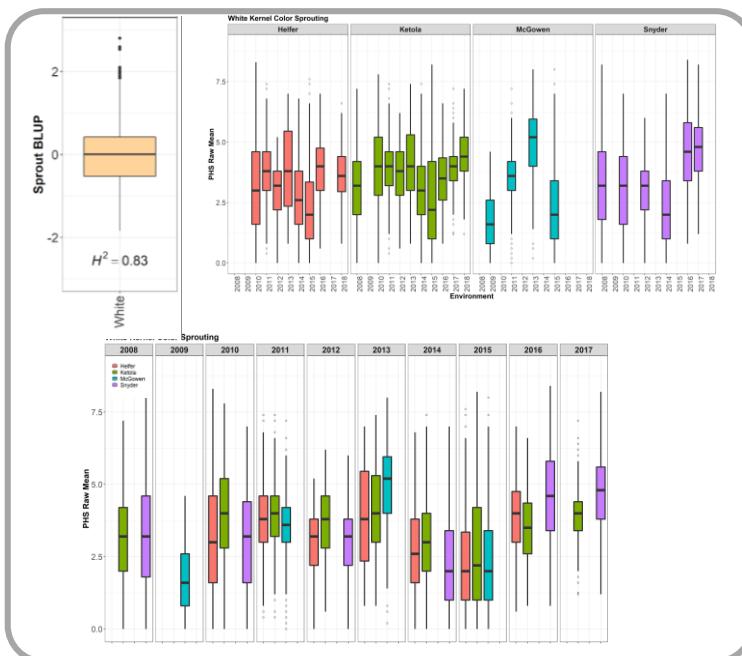
Detail

can be messy

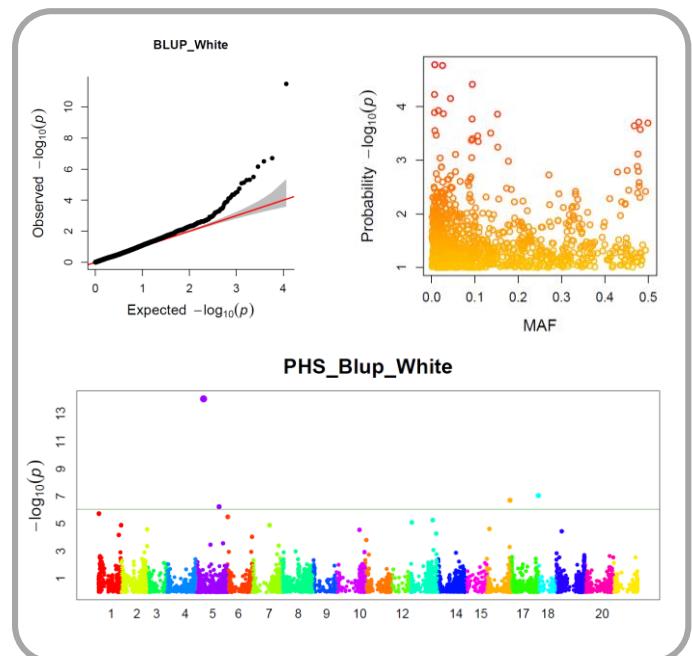
## Relationship



## Phenotype



## Significance



TIP #4 Use exploration to **better understand** your dataset before even thinking about perfecting the visualization



# Poster

The Point

varies

- i) open discussion on research
- ii) expose techniques to others
- iii) showcase the research

Use text to explain the main point without the author

Significance

TIP #6

Simplified notation to keep in line with story of the poster

Phenotype

TIP #7

Use grey to represent 'null' or 'without' to focus more on the story 'with'

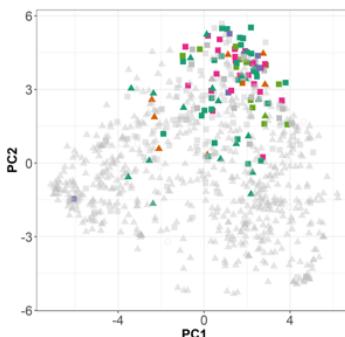
Relationship  
I focused on  
one many  
graphic that  
is the most  
telling

TIP #5

Accompanied  
by a visual of  
what the colors  
represent

## 2 GWAS

CORNELL MASTER NURSERY

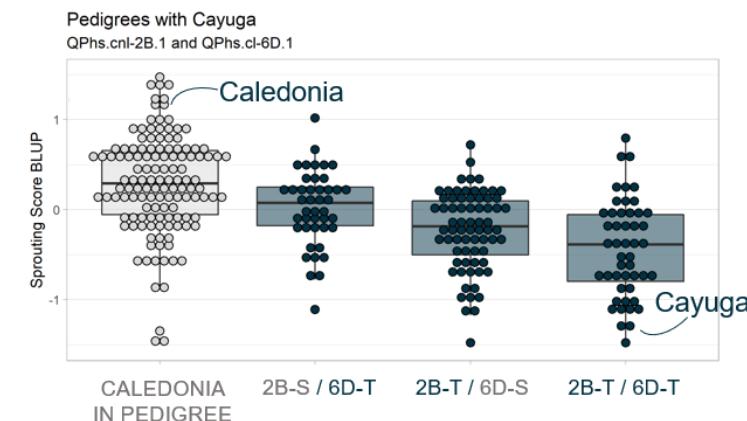
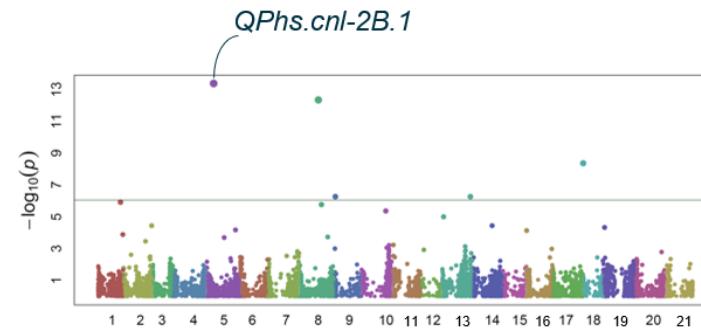


The PHS tolerant MTA QPhs.cnl-2B.1, contributed by the cultivar Cayuga, was highly significant at 174.8 Mb in the Cornell elite master nursery.

The other major PHS tolerant QTL contributed by Cayuga is on chromosome 6D, QPhs.cnl-6D.1, as shown in Munkvold et al., 2009

GBS-derived KASP markers have been created for both QTL

Pyramiding multiple Cayuga alleles resulted in increased PHS tolerance, however a single QTL present drastically reduces sprouting compared to Caledonia



# Presentation

The Point

always keep in mind the message of the graphic

Time

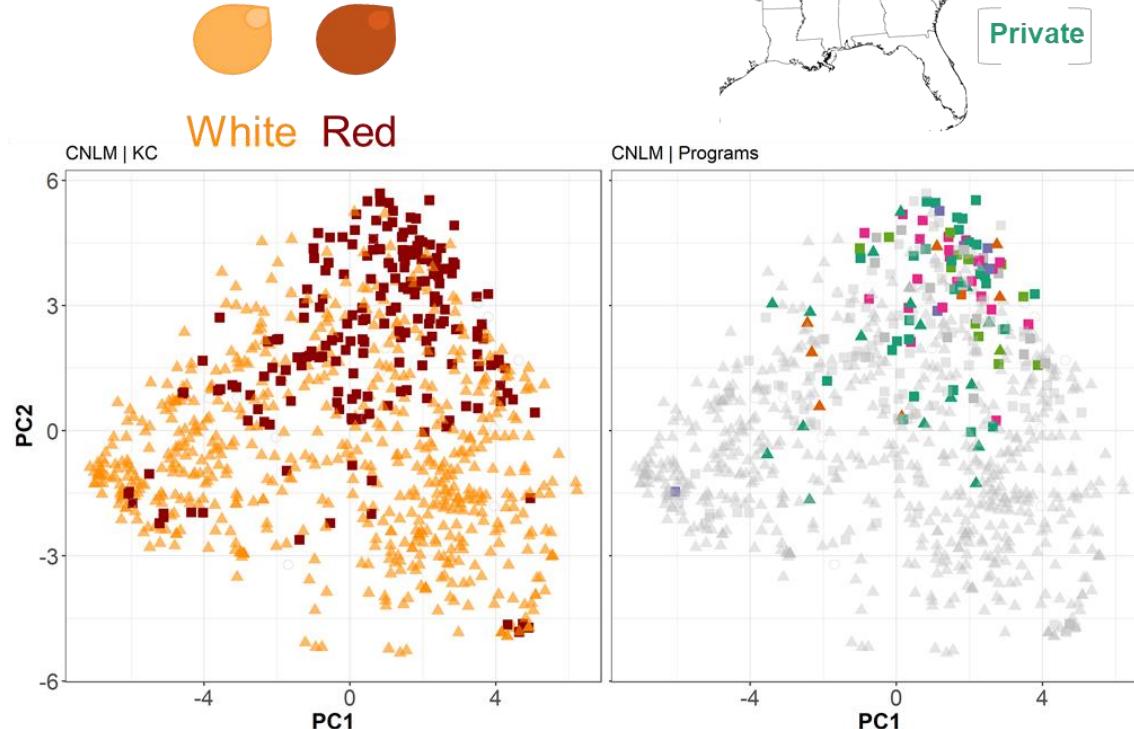
matter of seconds

Slide design was intentionally created with the message/conclusion in mind

TIP #10

Dual Purpose:  
Used two  
graphics to  
**introduce** my  
data in  
addition to  
portraying the  
relationship

## CNL Master Nursery



# Presentation

The Point

always keep in mind the message of the graphic

Time

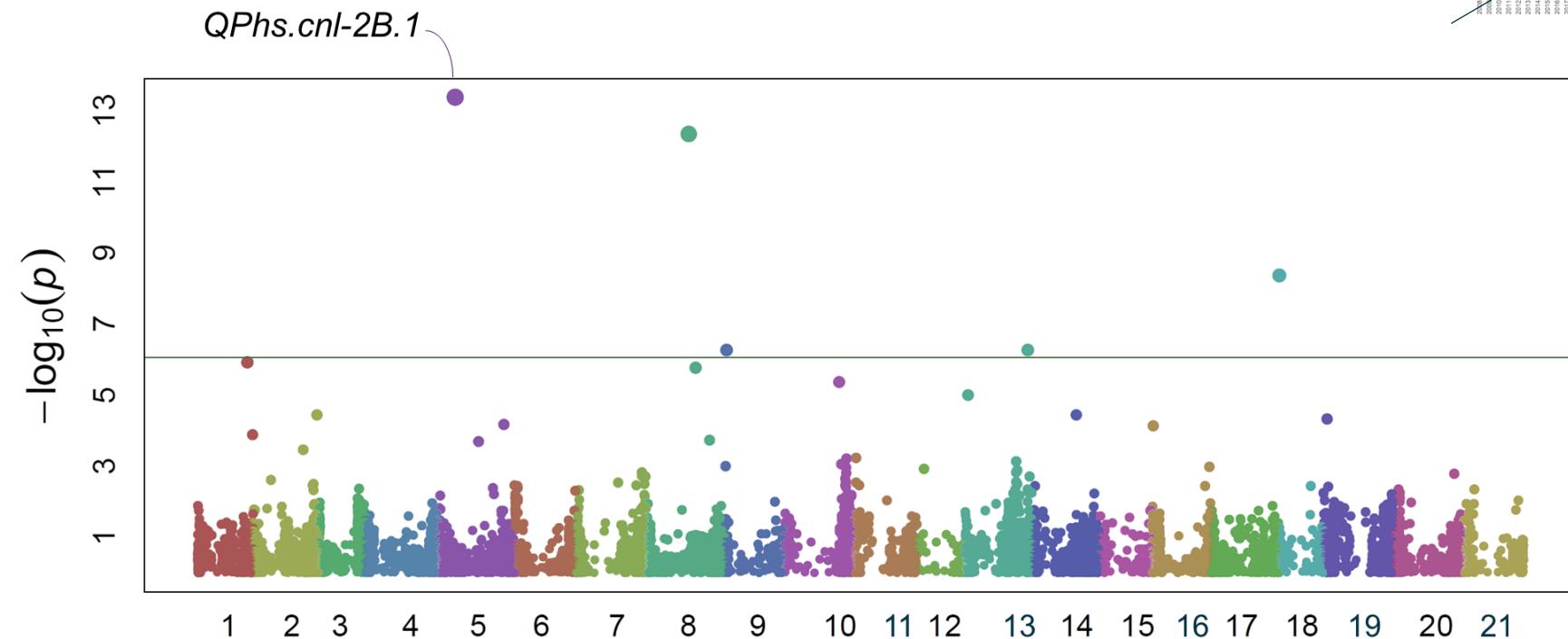
matter of seconds

Some 'basic' data/graphic (Phenotype) was omitted to save time

Significance

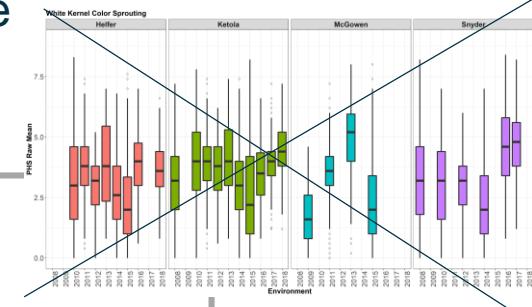
TIP #11

Use transitions to **focus** in on the main point



FarmCPU

Zanetti et al., 2000; Munkvold et al. 2009; Fofana et al. 2009; Zhang et al. 2013;  
Kumar et al. 2015; Martinez et al. 2018; Zuo et al., 2019



# Presentation

The Point

always keep in mind the message of the graphic

Time

matter of seconds

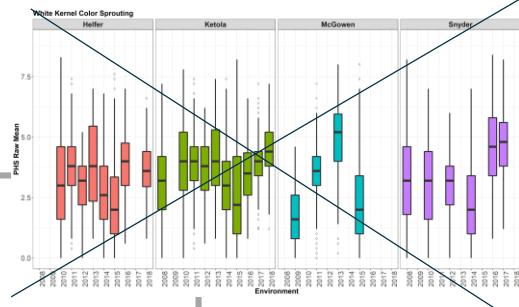
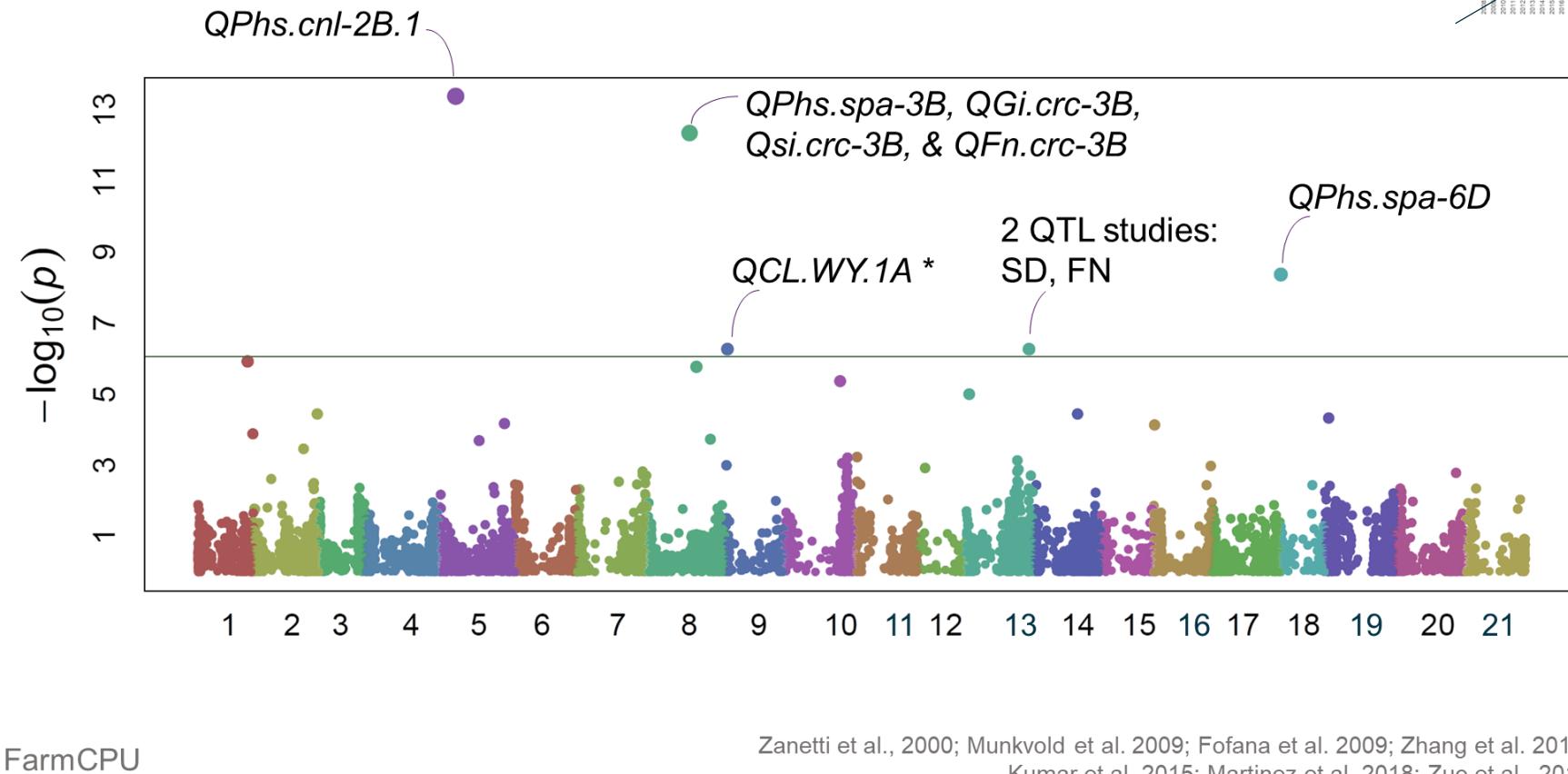
Some 'basic' data/graphic (Phenotype) was omitted to save time

Significance

TIP #11

Use transitions to **focus** in on the main point

Then briefly mention the other results to remain **thorough**



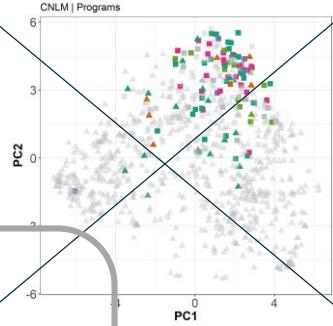
# Journal

The Point

detailed , thorough

## Relationship

Figures may be not included because  
they are referenced as text/citation

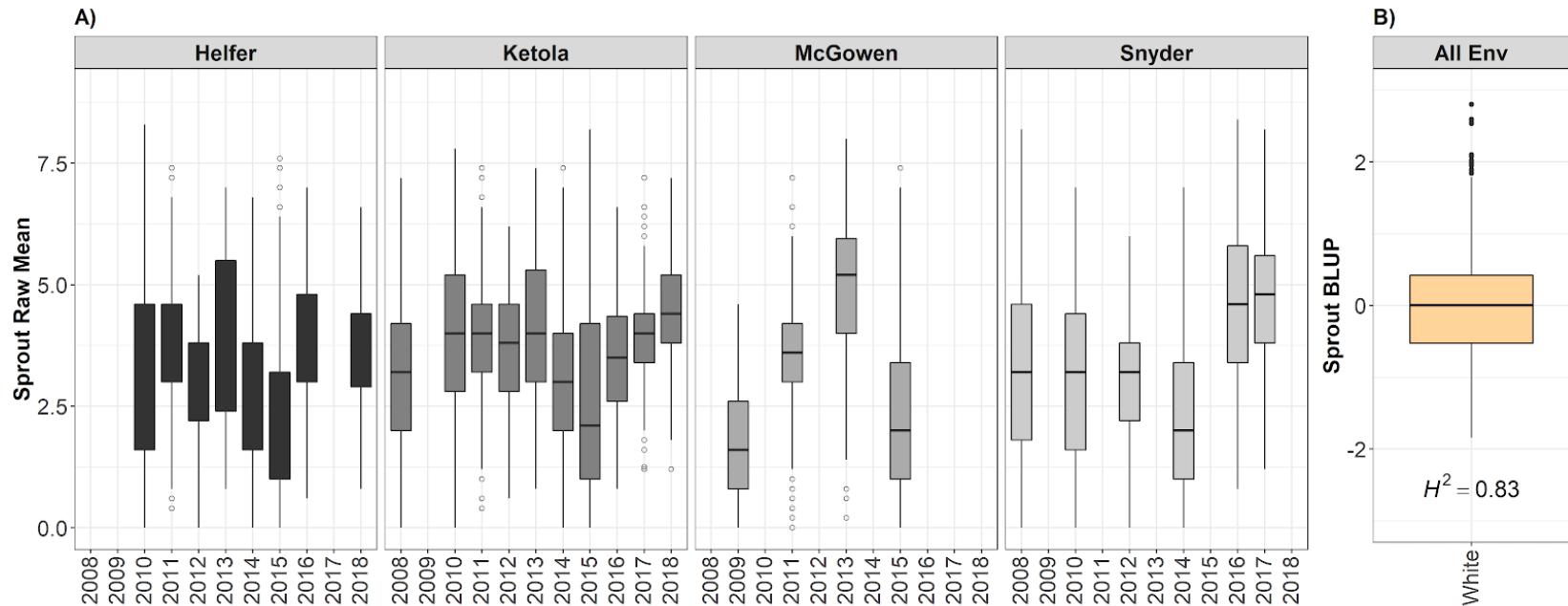


## Phenotype

TIP #12

Combined  
two variations  
of graphs into  
one to save  
space

In this case  
we are  
showing the  
raw  
phenotypic  
values and  
the estimated  
values



# Journal

The Point

detailed , thorough

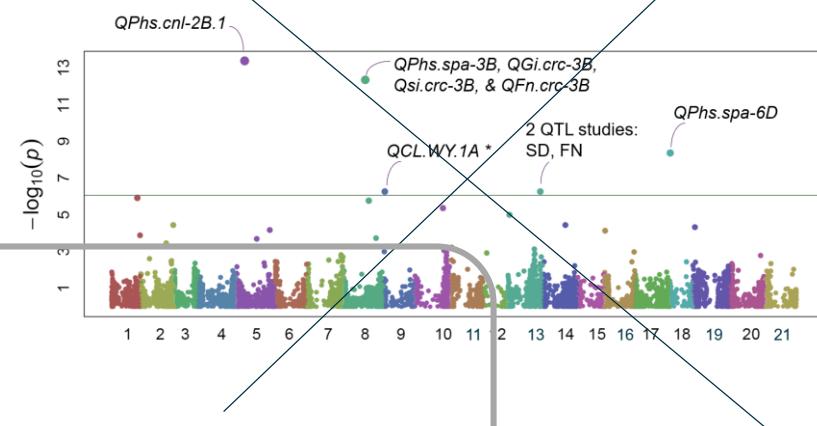
## Significance

TIP #13

The benefit of a table is the **amount of detail** you can communicate

**Table 2** - QTN associated with PHS traits.

QTL	chrom	pos	log10p	effect	KC	PC	Nearby QTL
QPhs.cn1-1B.2?	1B	640,605,809	6.51	0.5	Comb	0	QPHS.wsu-1B.2
QPhs.cn1-2B.3?	2B	163,977,776	6.31	-0.16	Comb	0	QPhs.spa-2B
QPhs.cn1-2B.1	2B	181,609,374	9.39	0.22	Comb	0	QPhs.cn1-2B.1
	2B	184,403,048	6.15	0.12	Comb	4	
QPhs.cn1-4B.1	4B	536,895,442	7.08	0.21	Comb	0	QPHS.wsu-4B.2
QPhs.cn1-5A.	5A	666,229,110	6.18	-0.43	Comb	0	Qfcgr.cas-5AL; Qgr.cas-5A



The benefit of a graphic is you can take in the magnitude of the significant **information quicker**

TIP #14

Before starting your Data Viz for a journal, read the publication requirements

**Note:** journals may be restrictive in regards to how you input tables.



# Public

The Point

the story

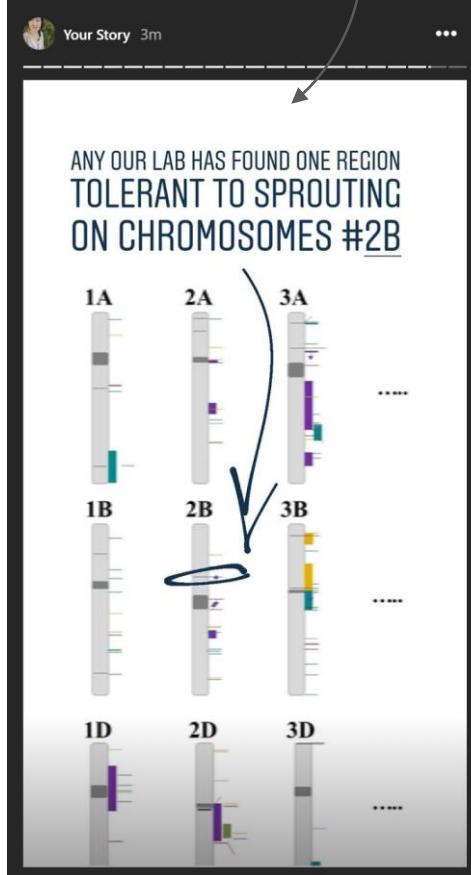
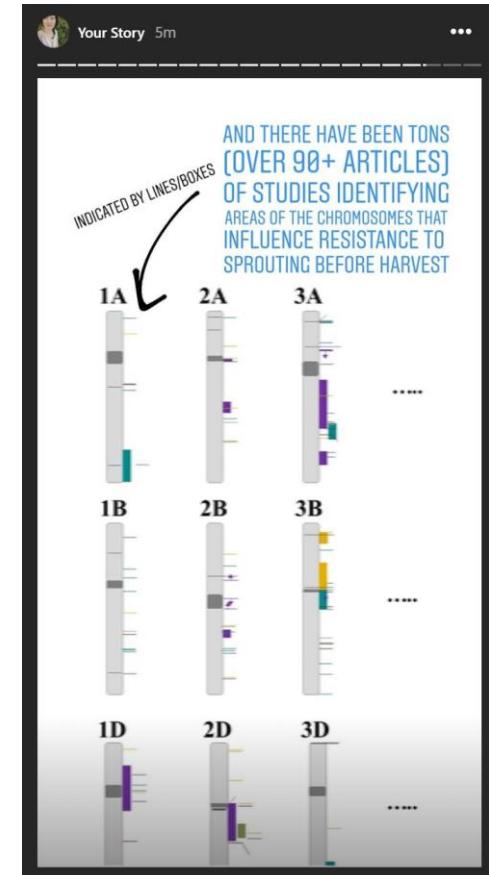
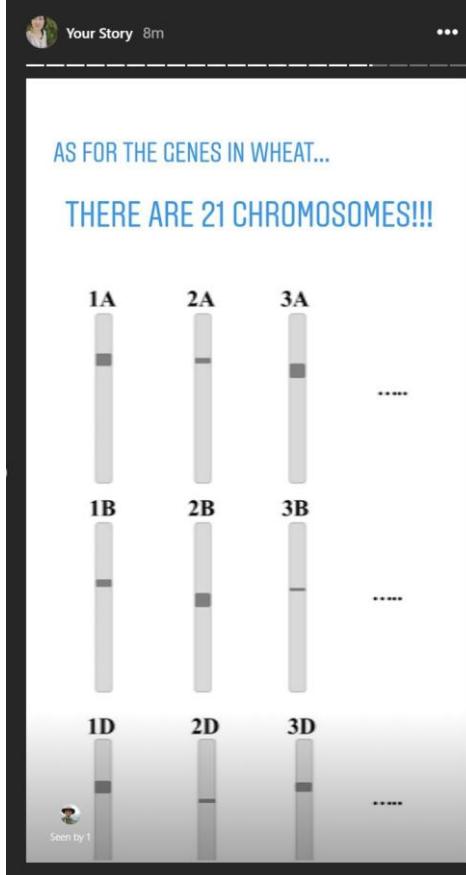
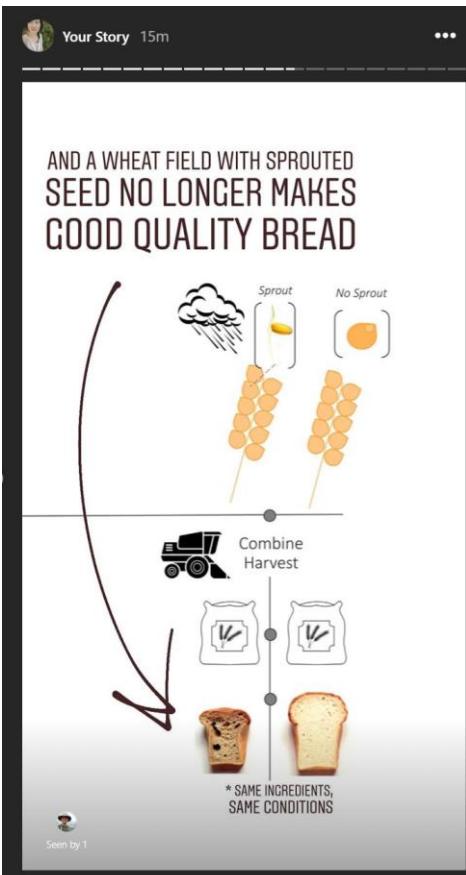
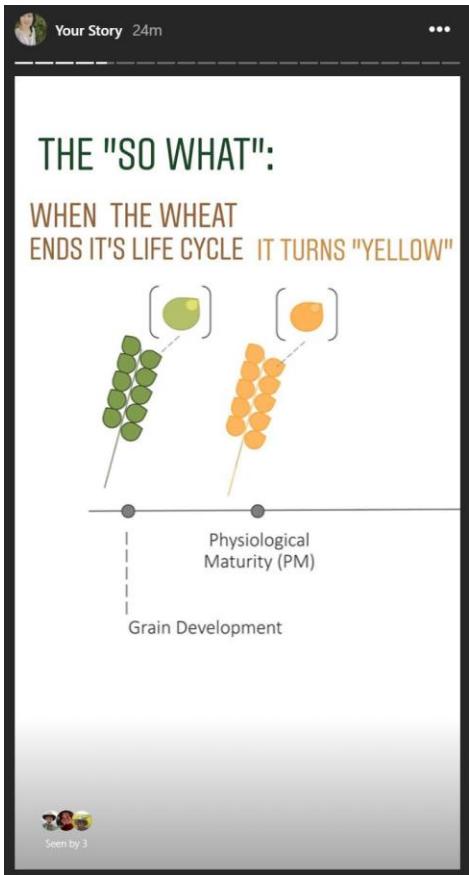
Time

instant

Detail

broad

Conclusion



The Story

Graphic

# Goals for today's webinar:

»»» Core Principles

»»» Examples from Others

»»» Examples from One Dataset

»»» Resources

Provide resources to improve your data visualizations

# Slide Deck (with resource links)

[shantel-martinez.github.io/DataViz2020](https://shantel-martinez.github.io/DataViz2020)

Screenshot me for later!

# How do I even read that graph?

To this day, I will see someone present a graphic that I've never seen before.

Resources like [The Data Visualization Catalogue](#) help me take my first step in recreating that graph

Provides a great **break down of different plot types**.

## The Data Visualisation Catalogue

About • Blog • Shop • Resources

CN 中文 ES Español RU Русский TR Türkçe

Search by Function

View by List



Arc Diagram



Area Graph



Bar Chart



Box & Whisker Plot



Brainstorm



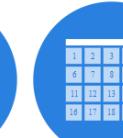
Bubble Chart



Bubble Map



Bullet Graph



Calendar



Candlestick Chart



Chord Diagram



Choropleth Map



Circle Packing



Connection Map



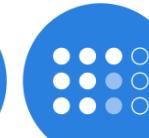
Density Plot



Donut Chart



Dot Map



Dot Matrix Chart



# What graphic do I present?

from Data to Viz

EXPLORE STORY ALL CAVEATS POSTER ABOUT CONTACT

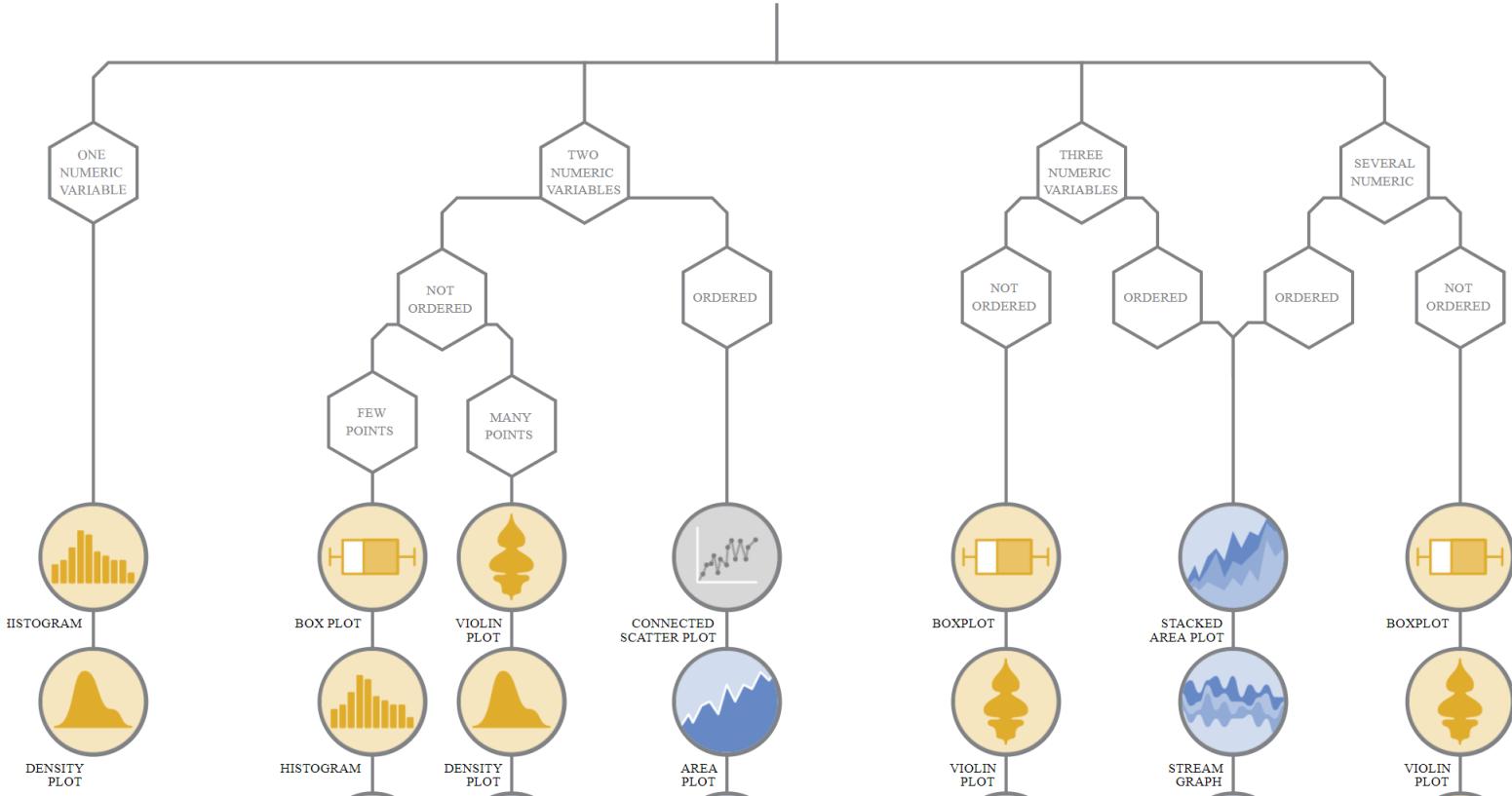
What kind of data do you have? Pick the main type using the buttons below. Then let the decision tree guide you toward your graphic possibilities.

Numeric Categoric Num & Cat Maps Network Time series

First, I always check the literature in the field for ideas

There are also **decision tree resources** that can help you make a decision on which graph to present for different types of data.

[LINK: Data to Viz](#)



# Books

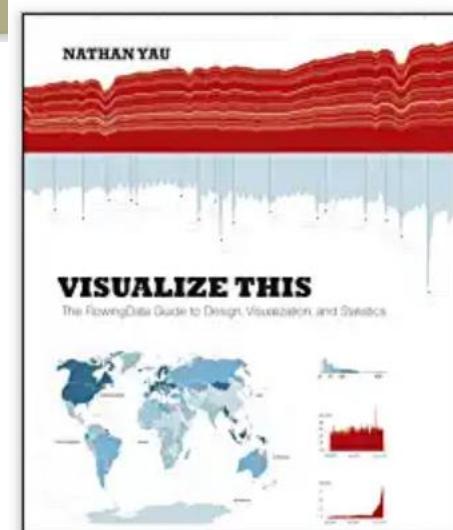
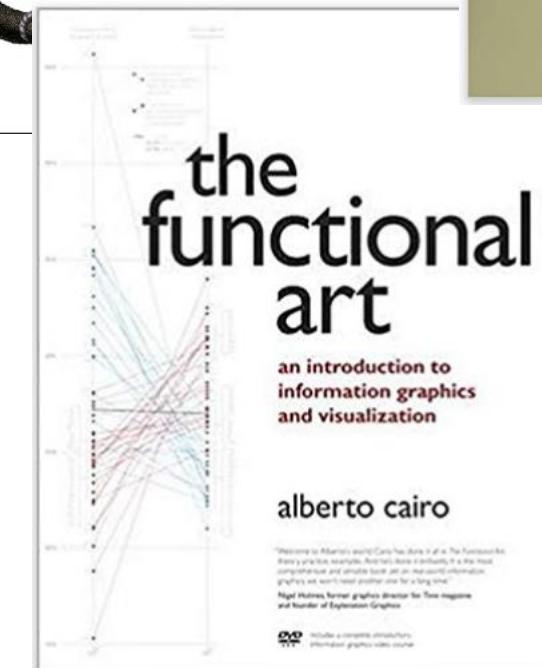
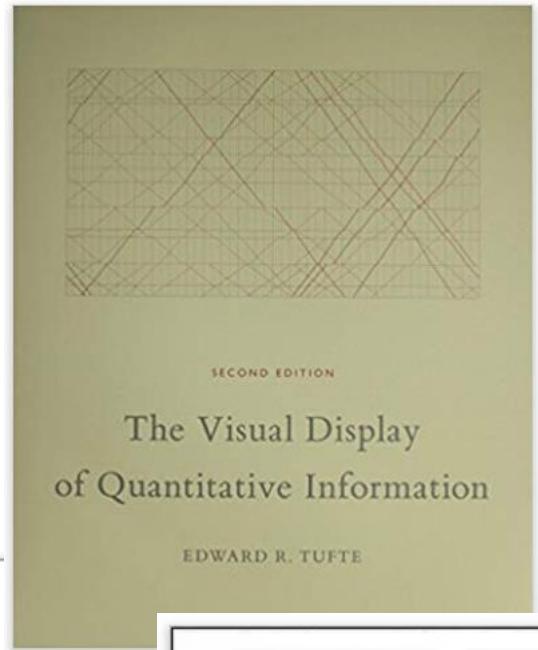
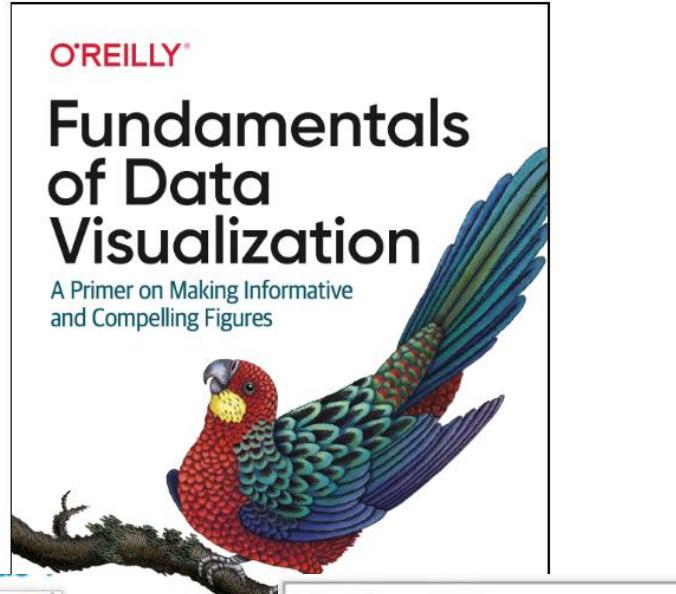
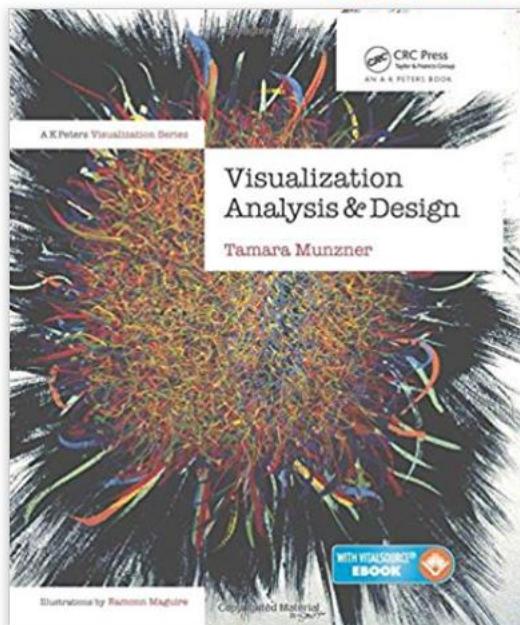
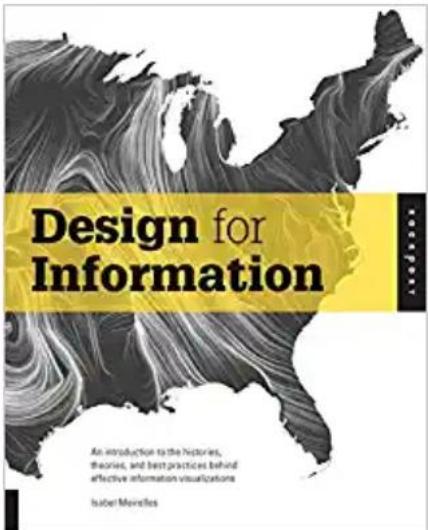
## Data Visualization Chapter 5

Free online

## In depth book reviews:

Free review

Must purchase or borrow books



# Scrolling



# Figure inspiration for R plots

# #TidyTuesday

# #tidyverse

# #dataviz

# So many people share code right along side of the graphic

**Follow data viz people, like Dr. Cédric Scherer, who give great examples of visualizations**



# Podcasts



## Link: Tidy Tuesday

**Episode 13: Christmas Eve 2019**  
Tidy Tuesday

Follow the show at @tidypod (<https://twitter.com/tidypod>) on Twitter! For show notes and to subscribe see [tidytuesday.com](http://tidytuesday.com) (<https://tidytuesday.com>)

DEC 24, 2019 • PLAYED

## Link: #DataTalk

**Data Storytelling: How to Make Your Data Visualizations More Effective w/...**  
DataTalk

We had a chance to talk with Nadieh Bremer about the steps to creating effective data visualizations to tell better stories. After graduating as a...

OCT 1, 2018 • 36 MINS

## Link: Data Stories

**140 | Data Visualization Society**  
Data Stories

We have the founding members of the Data Visualization society on the show to talk about how they got started.

MAY 28, 2019 • 5 MIN

**009 | Bridging academia and industry with Danyel Fisher**  
Data Stories

In this episode we talk about bridging academia and industry. We've touched upon this issue many times in the past that we decided to rec...

JUL 13, 2012 • 1 HR 16 MINS

## Link: Data Viz Today

**39: [Mini] 3 Design Tweaks that Make a Big Difference**  
Data Viz Today

I've been on a mission to improve my data visualization designs. In this episode I share three design tweaks that I've found to be...

FEB 12, 2019 • PLAYED

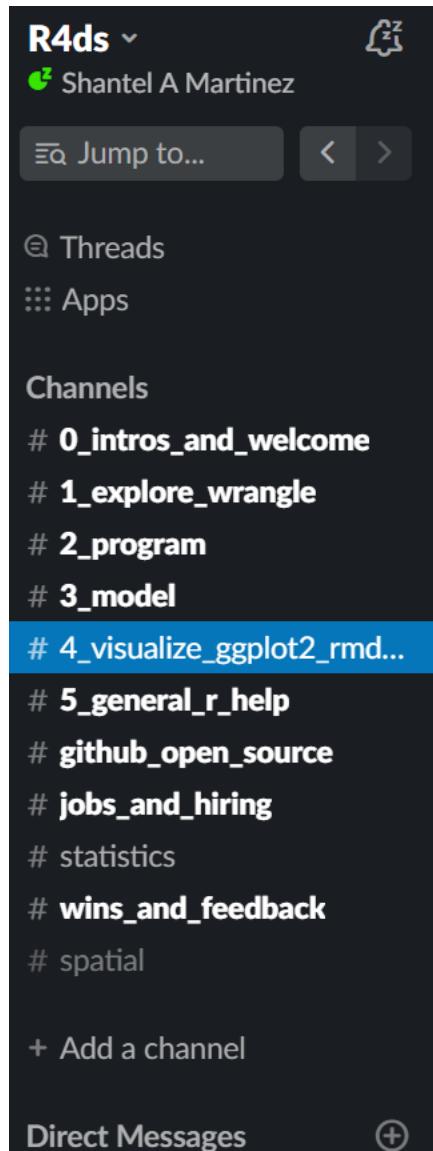
**50: How to Fill Your Data Viz Toolbox — a New Years Resolution that won't dr...**  
Data Viz Today

Happy new year! It's time for goal-setting, right? One small thing you can do throughout the year that will make you a better information des...

DEC 31, 2019 • 21 MINS

# Discussion

[Link: R for Data Scientists](#)



A screenshot of the R4ds Slack interface. At the top, there's a header with "R4ds" and a user icon for Shantel A Martinez. Below the header are buttons for "Jump to..." and navigation arrows. On the left, there are sections for "Threads" and "Apps". The main area lists several Slack channels, each preceded by a "#": #0\_intro\_and\_welcome, #1\_explore\_wrangle, #2\_program, #3\_model, #4\_visualize\_ggplot2\_rmd..., #5\_general\_r\_help, #github\_open\_source, #jobs\_and\_hiring, #statistics, #wins\_and\_feedback, and #spatial. The channel "#4\_visualize\_ggplot2\_rmd..." is highlighted with a blue background. At the bottom, there's a button for "+ Add a channel" and a "Direct Messages" link.



Slack is a well organized “discussion forum”

R4DS slack has a lot of R resources,  
such as the ggplot2 channel



# Discussion

[Link: R for Data Scientists](#)

R4ds v Shantel A Martinez

Jump to... < >

Threads Apps

Channels

# 0\_intro\_and\_welcome  
# 1\_explore\_wrangle  
# 2\_program  
# 3\_model  
**# 4\_visualize\_ggplot2\_rmd...**  
# 5\_general\_r\_help  
# github\_open\_source  
# jobs\_and\_hiring  
# statistics  
# wins\_and\_feedback  
# spatial

+ Add a channel

Direct Messages +

#4\_visualize\_ggplot2\_rmd\_etc

3,848 | 1 | Get help with communicating in R.

December 5th, 2019

Bongyeol Y 5:10 PM

Would you like to show me how to get rid of all of the legends, or one of them from the graph below? Thank you!

```
mtcars %>%  
  ggplot(aes(  
    x = mpg,  
    y = disp,  
    shape = factor(cyl),  
    color = disp > mean(disp))  
  ) +  
  geom_point(size = 5) +  
  theme_classic(base_size = 14)
```

example.png

The scatter plot displays the relationship between engine displacement (disp) and fuel efficiency (mpg). The x-axis represents mpg, ranging from 10 to 35. The y-axis represents disp, ranging from 100 to 450. The data points are categorized by cylinder count (4, 6, or 8) and whether their displacement is above (TRUE) or below (FALSE) the mean displacement. The plot features a legend for each category, with circles for 4 cylinders, triangles for 6 cylinders, and squares for 8 cylinders. Points where disp > mean(disp) are colored teal, while points where disp <= mean(disp) are colored red. The overall aesthetic is clean, using a white background and a classic font style.



# Wiki pages

r-statistics.co by Selva Prabhakaran

## Top 50 ggplot2 Visualizations - The Master List (With Full R Code)

What type of visualization to use for what sort of problem? This tutorial helps you choose the right type of chart for your specific objectives and how to implement it in R using ggplot2.

This is part 3 of a three part tutorial on ggplot2, an aesthetically pleasing (and very popular) graphics framework in R. This tutorial is primarily geared towards those having some basic knowledge of the R programming language and want to make complex and nice looking charts with R ggplot2.

- [Part 1: Introduction to ggplot2](#), covers the basic knowledge about constructing simple ggplots and modifying the components and aesthetics.
- [Part 2: Customizing the Look and Feel](#), is about more advanced customization like manipulating legend, annotations, multiplots with faceting and custom layouts
- [Part 3: Top 50 ggplot2 Visualizations - The Master List](#), applies what was learnt in part 1 and 2 to construct other types of ggplots such as bar charts, boxplots etc.

[Link: Top 50 ggplot2](#)

### Correlogram

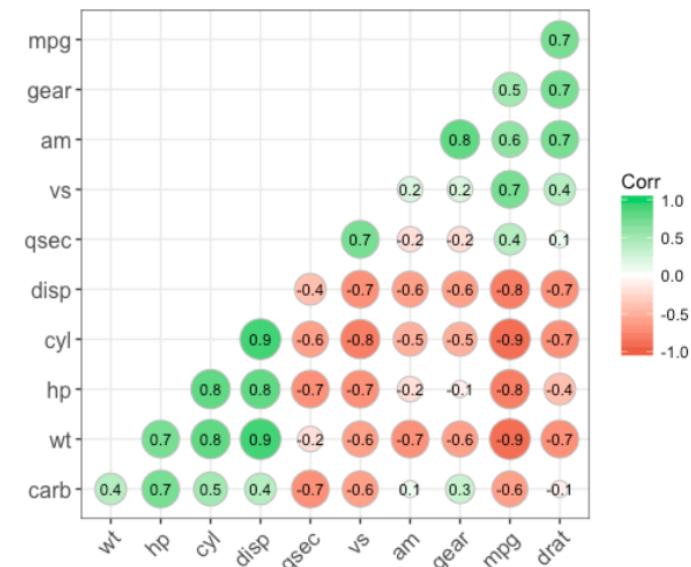
Correlogram let's you examine the corellation of multiple continuous variables present in the same dataframe. This is conveniently implemented using the `ggcorrplot` package.

```
# devtools::install_github("kassambara/ggcorrplot")
library(ggplot2)
library(ggcorrplot)

# Correlation matrix
data(mtcars)
corr <- round(cor(mtcars), 1)

# Plot
ggcorrplot(corr, hc.order = TRUE,
           type = "lower",
           lab = TRUE,
           lab_size = 3,
           method="circle",
           colors = c("tomato2", "white", "springgreen3"),
           title="Correlogram of mtcars",
           ggtheme=theme_bw)
```

Correlogram of mtcars



# Blogs

[Link: Nightingale](#)

## Nightingale

The Journal of the Data Visualization Society



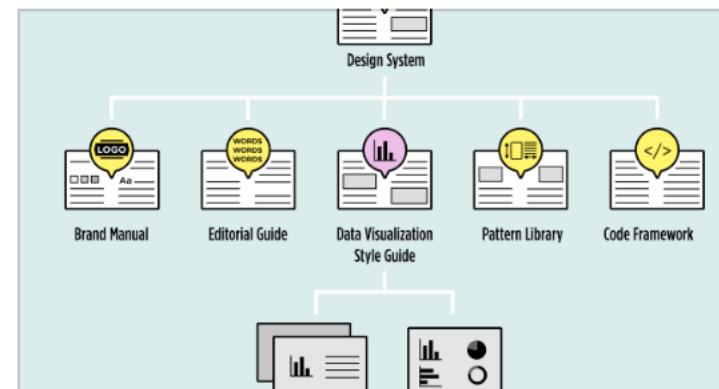
### Choosing the Right Tools for Data Visualization

A conversation about favourites, how to approach learning a new tool, and why data sketching could change your life



Duncan Geere

Dec 10, 2019 · 8 min read



### What Are Data Visualization Style Guidelines?

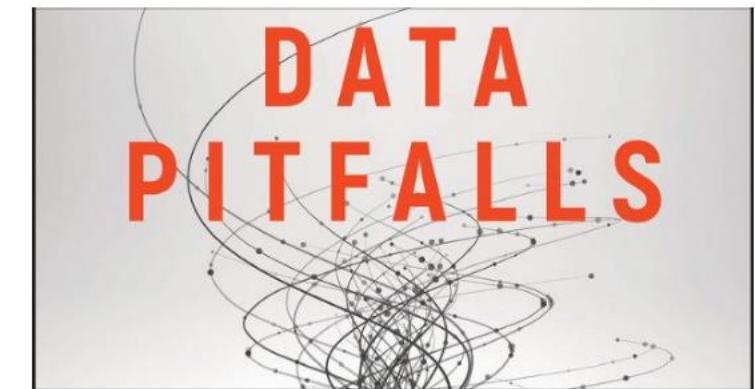
Data visualization style guides are standards for formatting and designing representations of information.



Amy Cesal

Jul 10, 2019 · 8 min read

[LINK](#)



### 'Avoiding Data Pitfalls'—an Interview With Ben Jones

The founder and CEO of Data Literacy's new book will help you avoid common data analysis and visualization mistakes



Neil Richards

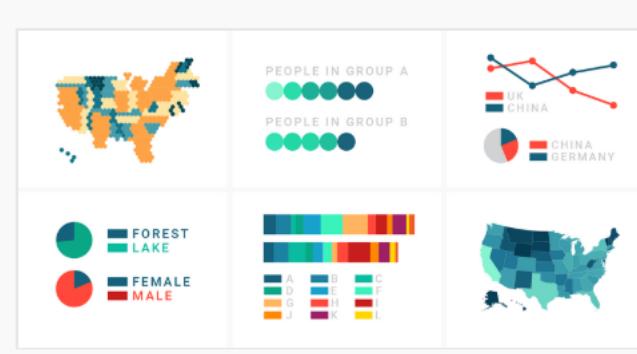
Dec 23, 2019 · 10 min read ★

# Blogs

## CHARTABLE

A blog by [Datawrapper](#)

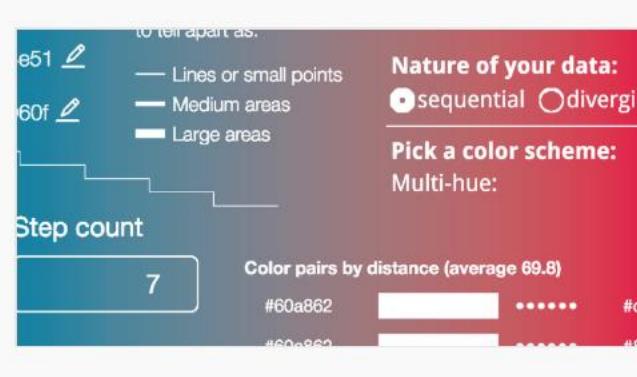
[LINK](#)



Thoughts & How To's / May 29, 2018

## What to consider when choosing colors for data visualization

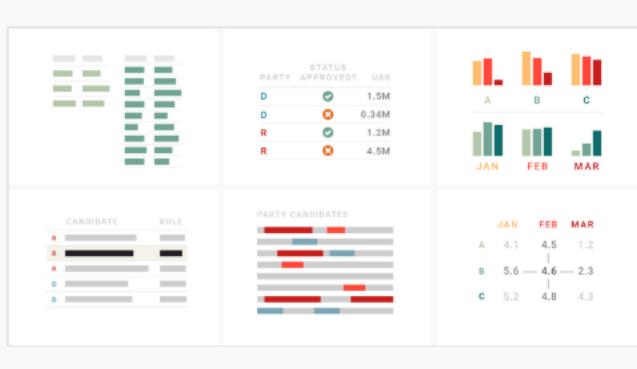
[LINK](#)



Thoughts & How To's / Jul 31, 2018

## Your Friendly Guide to Colors in Data Visualisation

[LINK](#)



Thoughts & How To's / May 21, 2019

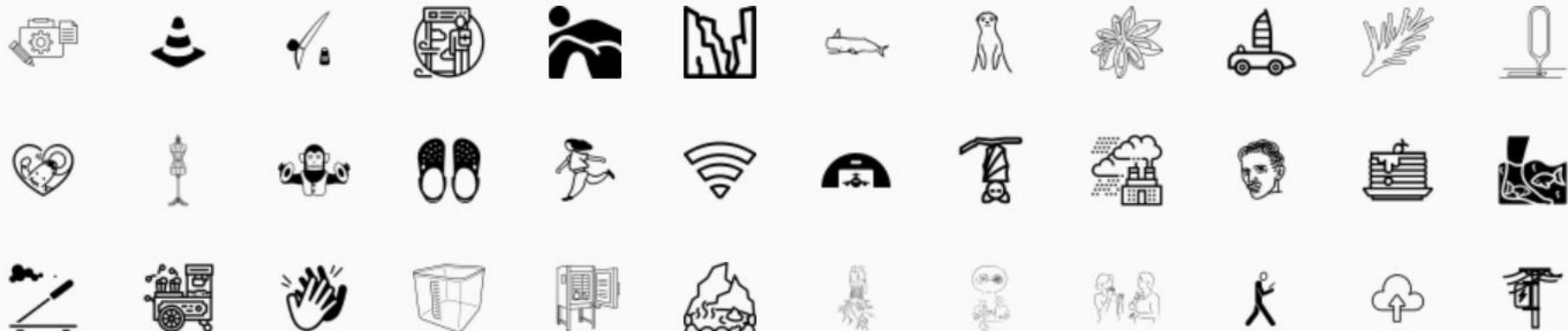
## What to consider when creating tables



# Icons for everything

Over 2 Million curated icons, created by a global community

Search for anything



[Keep Browsing](#)

Creative community  
Free: give credit to artists  
Subscription: Edit colors



[Link: Data Visualization Society](#)



## DATA VISUALIZATION SOCIETY

### NEWSLETTER AND SLACK

Occasionally receive updates conveniently located in your inbox.

Opt in to joining the Slack chat platform to communicate with fellow practitioners about data visualization topics like design, optimization and development when you become a member.

BECOME A MEMBER



[Link: Data Visualization Society](#)

A screenshot of a Slack interface. At the top left, it says "Data Visualiza..." with a dropdown arrow. Next to it is a user icon for "Shantel A Martinez". Below the header are buttons for "Jump to..." and navigation arrows. On the left, there's a sidebar with "Threads" and "Apps" buttons, followed by a list of channels: "# -events", "# -introductions", "# -topics-in-data-viz", "# dvs-suggestions", "# general", "# help-general", "# help-slack", "# share-critique" (which is highlighted with a blue background), "# share-inspiration", "# share-showcase", and "# -announcements". A small red circle with the number "1" is at the bottom right of the sidebar.

Data Viz Society slack discussions are not isolated to just R.  
More about the end product: graphic

A great resource for graphical inspiration.

# Thank you for your attention

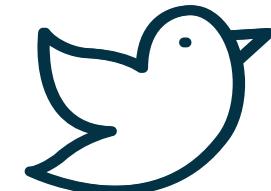
Email



Slide Deck (with links)



Twitter



shantel.a.martinez@gmail.com

shantel-martinez.github.io/DataViz2020

@s\_amealia



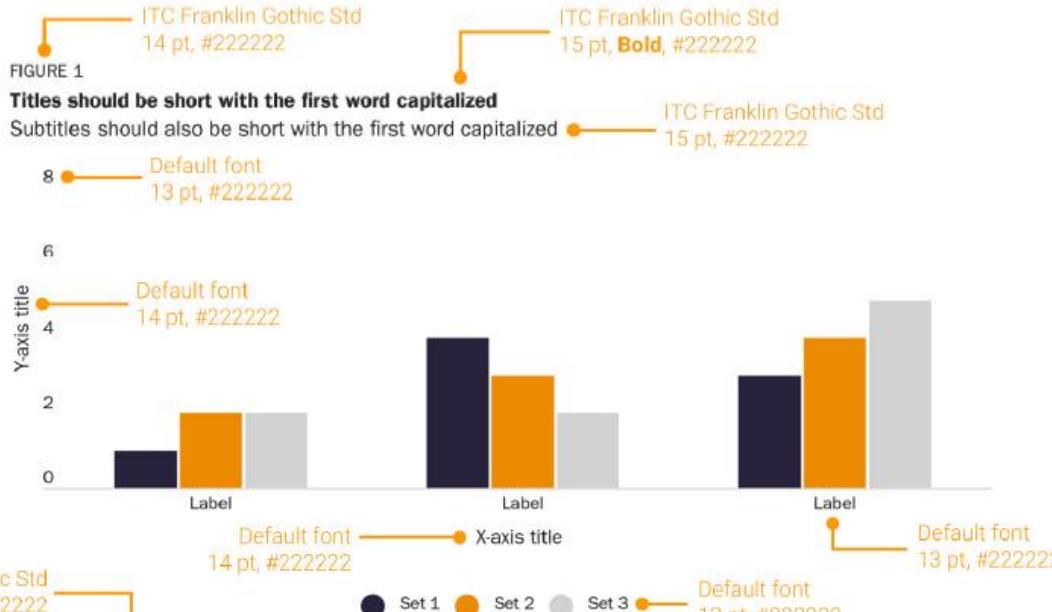
# The Data Visualization Webinar

In time, you will receive an email with a short survey and the link to these slides  
In a couple weeks you will also receive the recording of today's webinar

Brought to you by the ASA, CSSA, and SSSA  
Graduate Student Committee

# Style Guides

## Link: The Cato Institute



\*We have omitted the Legal Gender indicator from the calculations used in this figure because it is the only indicator in the HFI that appears in only two years (2015 and 2016).

**Source:** James Gwartney et al., Economic Freedom of the World: 2018 Annual Report (Vancouver: Fraser Institute, 2018).

ITC Franklin Gothic Std  
12 pt, #222222

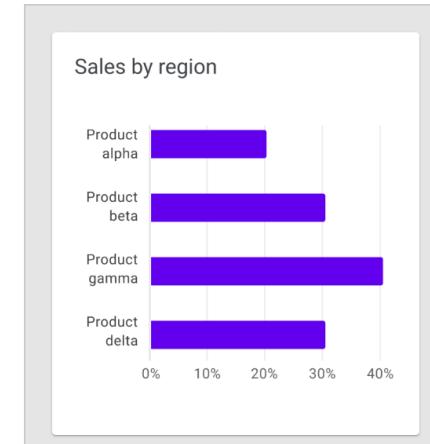
## Link: Material Design

### Text orientation

Text labels should be placed horizontally on the chart so that they are easy to read.

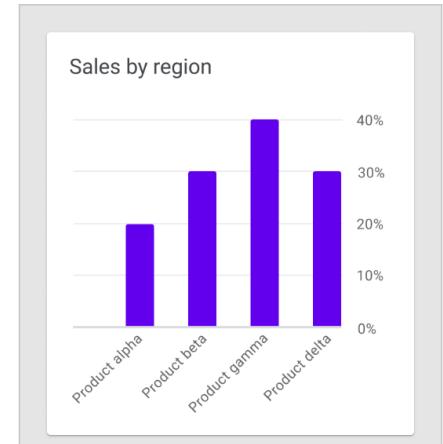
Text labels should not:

- Be rotated
- Stacked vertically



#### Do.

Orient text horizontally on bar charts, rotating the bars if needed to make space.



#### Caution.

Don't rotate bar labels, as it makes them difficult to read.