MSA 8020
Data Visualization

Yichen Cheng

Georgia State University | J. Mack Robinson College of Business

Robinson

Software Requirement

RStudio (Majority of the class)

- You need to install both R and Rstudio
- Rstudio is like a more user-friendly version of R

Tableau (Week 2)

- https://www.tableau.com/products/desktop/download
- Please make sure to install it before class 2.

Useful websites

iCollege:

- codes, lecture notes, homework/project submission
- https://icollege.gsu.edu/

Piazza:

- discussion:
- piazza.com/gsu/fall2020/msa8020

TAs and Office Hours

Nimeelitha Akkiraju: nakkiraju1@student.gsu.edu

Andrea Ekey: aekey1@student.gsu.edu

Jiahui Li: jli69@student.gsu.edu

Serkan Comu: scomu1@student.gsu.edu

Answering questions on Piazza: during lecture times.

TA office hours (for coding related questions):

- Wednesday: 2:00-3:00 pm
- Friday: 10:30-11:30 am

My Office Hours:

Monday: 4:00-5:00 pm or by appointment

Email: ycheng11@gsu.edu

Zoom link: https://zoom.us/j/9837544147

Syllabus

7 sessions

- First six sessions are lecture based.
- Last session will be final project presentation.

Each lecture

- First part: background and concept
- Second part: code

Syllabus

NO required textbook.

A list of recommended resources:

Syllabus

Class outline:

- Introduction and data preparation
- Tableau
- Basic R plot
- R ggplot
- Interactive plot
- Spatial data

Syllabus

Form group in week 1.

- Each group 3-5 people
- Once a group is formed, post: group members' names, group leader, group name on piazza
- You are encouraged to be in the same group throughout the course

Syllabus

There are three homework assignments.

- Each team submits one copy on iCollege.
- One week to finish the homework, due at 6 PM Tuesday.
- Submit: codes, slides (at most 10 pages) and a written report (at most 5 pages).

Final project (Week 7):

- 15 minutes presentation.
- Due Sunday midnight of week 7.
- Submit: codes, slides (at most 15 pages), a written report (at most 10 pages), and a team member evaluation.

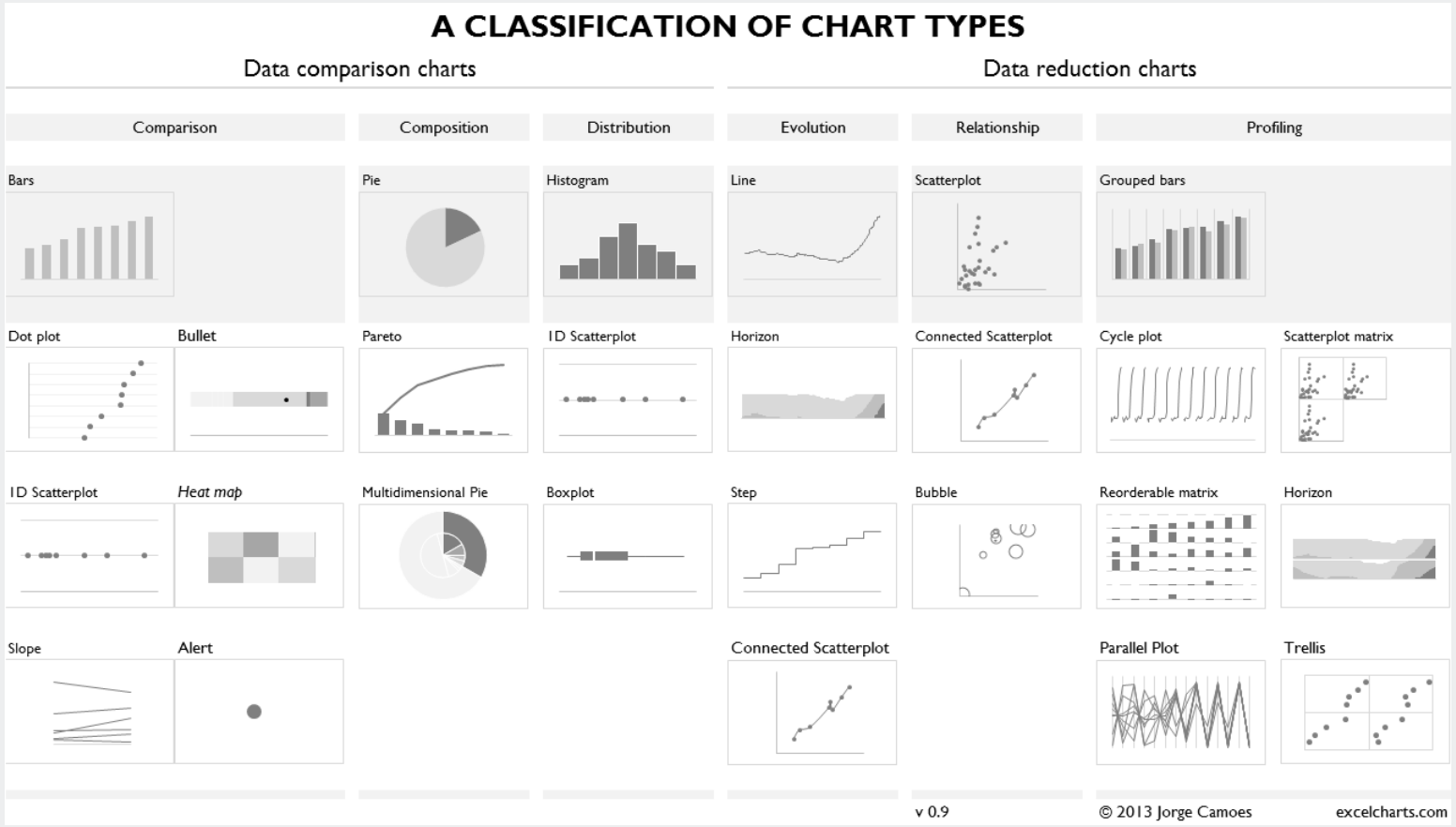
Syllabus

Some useful data websites:

- https://www.kaggle.com/datasets
- https://github.com/awesomedata/awesome-public-datasets
- http://archive.ics.uci.edu/ml/index.php
- https://www.data.gov/
- https://opendata.socrata.com/

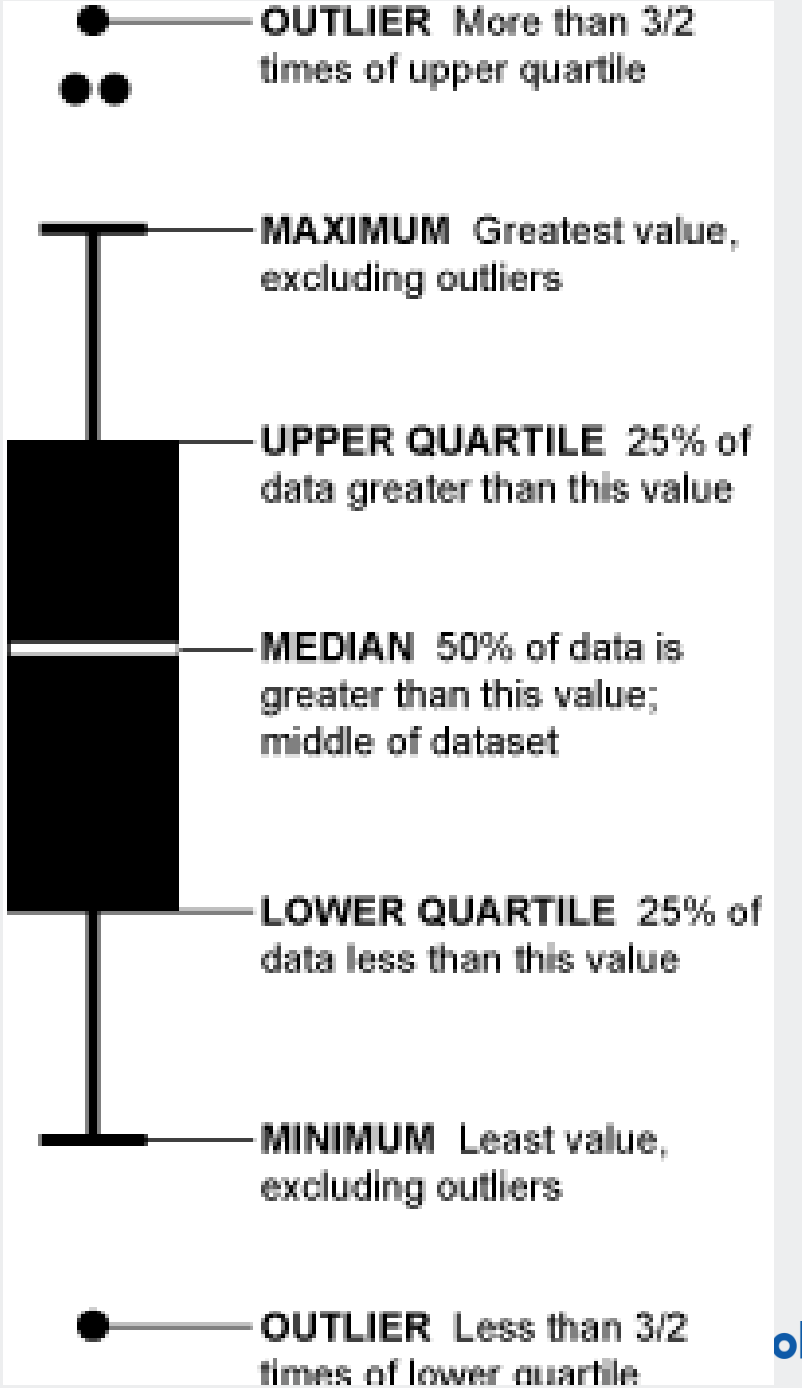
Why data visualization?

Different chart types



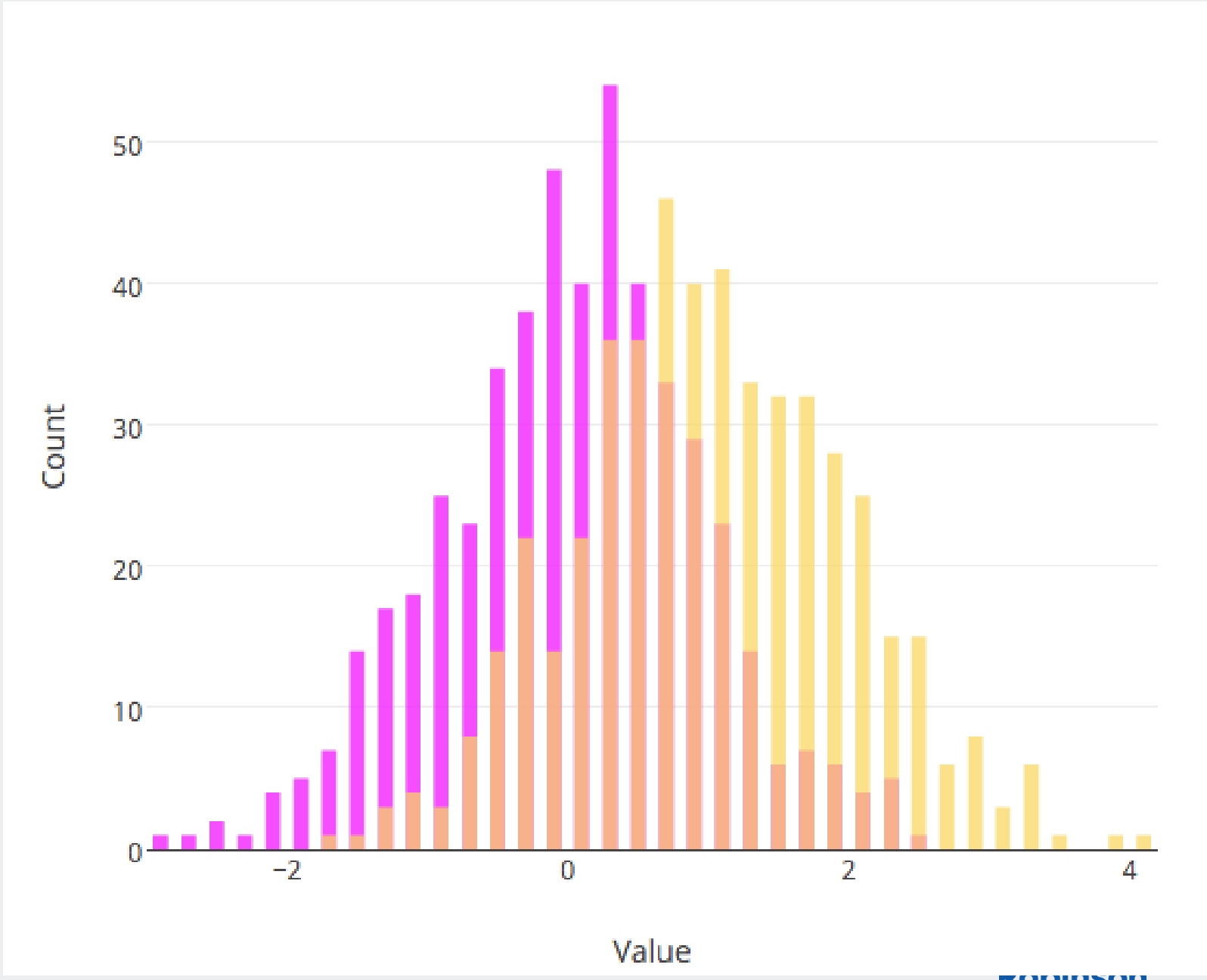
Boxplot

Effective for understanding your data's distribution



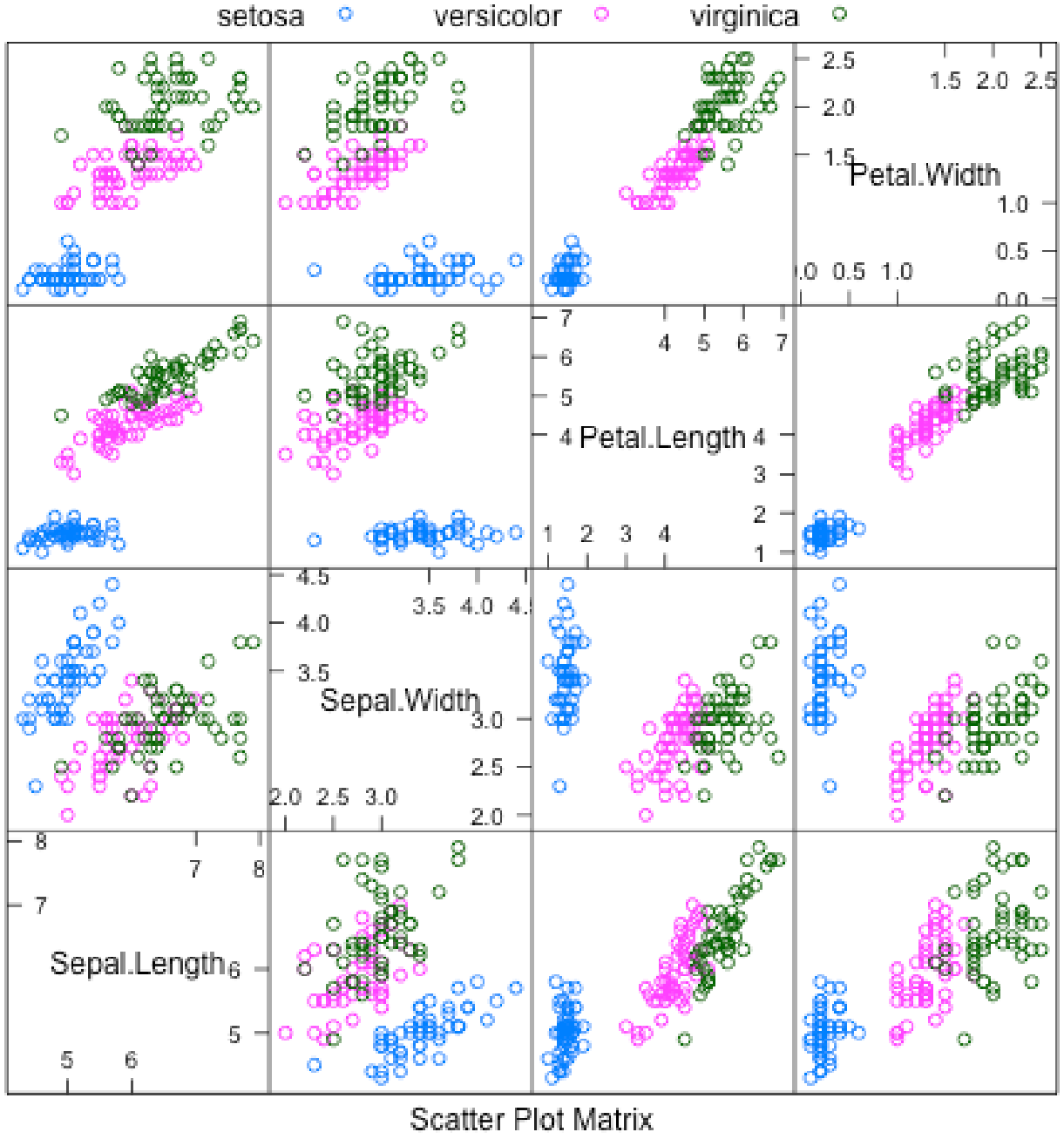
Histogram

Effective for understanding data's distribution, in light of counts.



Scatterplot

Effective at data comparison.



Different data types and possible Plots

https://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/DataPresentation/DataPresentation7.html

https://www.r-graph-gallery.com/base-R.html

https://www.r-graph-gallery.com/ggplot2-package.html

R introduction

- Install/call packges
- Read in data set
- View, summary
- Subset data
- Create new variables

R introduction – Basic Concept

Package:

- Install a package: buy a toolbox (a collection of tools/functions)
- Library/require a package: take out a toolbox and bring it on hand

Function:

- Function: tool
- Usual form: function_name(input) -> output
- Example: sqrt(4) -> 2

~~What chart should I use?~~

What am I trying to do?

http://www.nytimes.com/interactive/2016/09/13/us/politics/what-separates-voters-and-nonvoters.html?_r=0

Hans Rosling Video

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

Secondary Audience

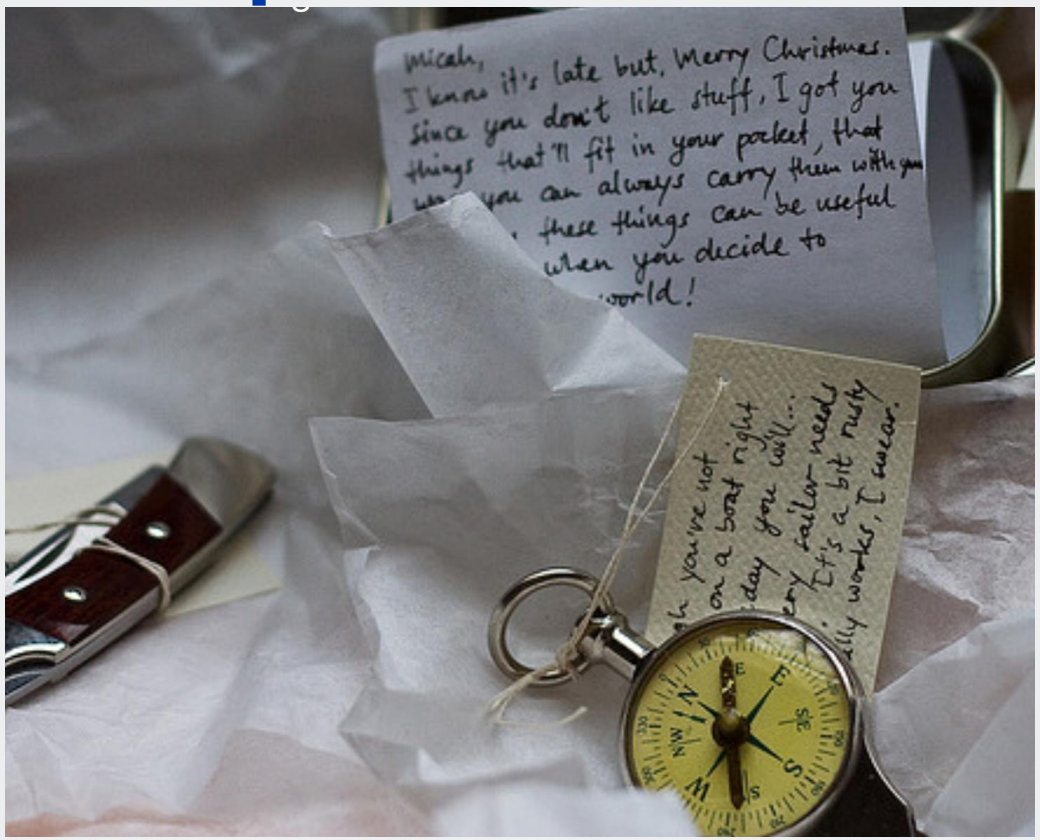
Sometimes the people not in the room matter more than those engaged in the presentation.

Remember to present to these secondary audiences.

Principle #1: Explanation before Information

"You take the blue pill, the story ends. You wake up in your bed and believe whatever you want to believe. You take the red pill, you stay in wonderland, and I show you how deep the rabbit hole goes."

Principle # 2: Use Text & Annotations



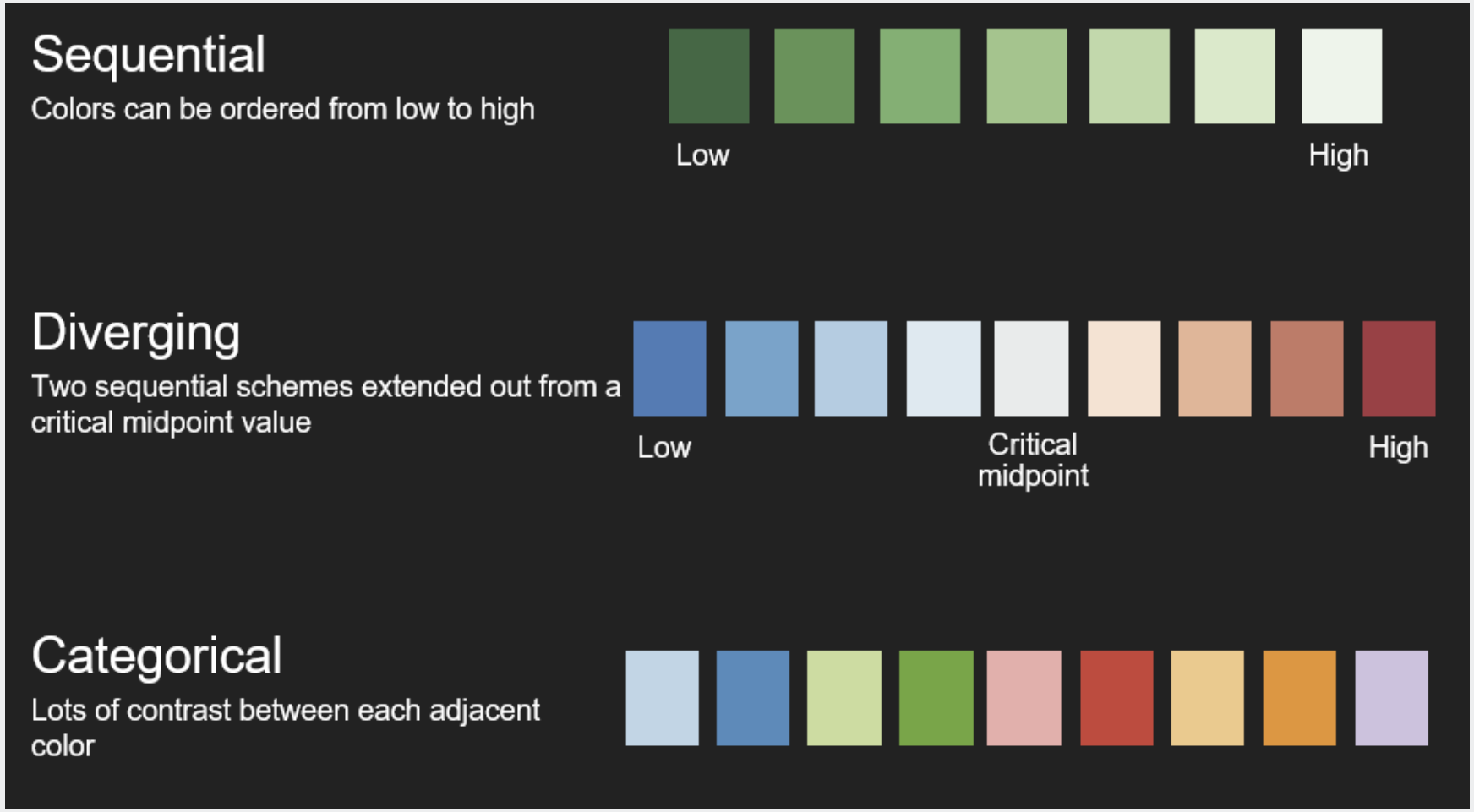
Keep text readable

Prioritize what you want them to read - Title vs. a legend vs a note

http://flowingdata.com/2013/10/22/working-with-text-in-r/

http://flowingdata.com/2016/07/06/annotating-charts-in-r/

Principle #3: Use Color Purposefully



r color cheatsheet:
https://www.nceas.ucsb.edu/~frazier/RSpatialGuides/colorPaletteCheatsheet.pdf