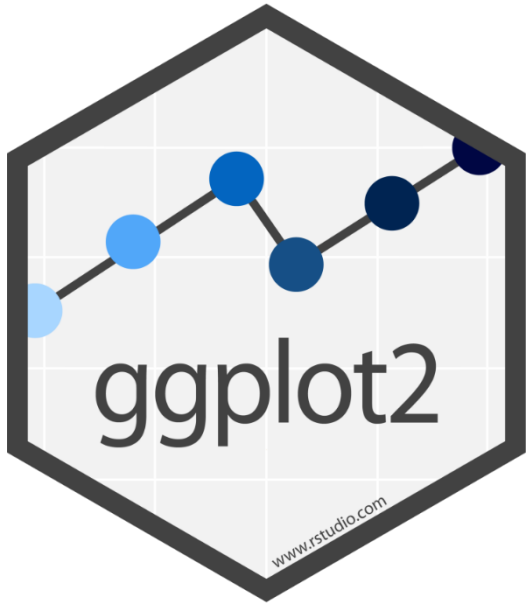


Introduction to Data Visualization in R

Instructor: **Pete Lawson**

JOHNS HOPKINS
LIBRARIES
Data Services

Website: dataservices.library.jhu.edu

Email: dataservices@jhu.edu

JHU Data Archive: archive.data.jhu.edu

Using Zoom

- Mute your audio
- Turn off your camera
- To ask questions:
  - Raise your hand – we'll be with you shortly
  - Write in the public chat or private chat a TA directly
  - Unmute your mic and speak up!
- During hands-on activities:
  - Mark YES or NO to let us know if you're having trouble
  - Write in the public chat

About this webinar

- This webinar will be recorded
- This webinar consists of a mix of lectures and hands-on activities
- After the workshop, you'll receive a complete version of all R files and workshop materials
- If you have any questions after the workshop, email us at dataservices@jhu.edu

JHU DATA SERVICES

HELPING YOU NAVIGATE DATA

WE HELP FACULTY, RESEARCHERS AND STUDENTS

FIND

USE

MANAGE

VISUALIZE

SHARE

FIND OUT MORE

GO TO dataservices.library.jhu.edu

EMAIL dataservices@jhu.edu

SHARE AT archive.data.jhu.edu

JOHNS HOPKINS LIBRARIES | Data Services

This workshop will

1. Introduce basic concepts of data visualization
2. Introduce you to the terminology and syntax for base R and ggplot2
3. Walk you through creating some basic plots and graphs using base R and ggplot2

What we WON'T cover

Introductory R concepts

- Basic R syntax and terminology
- Installing and loading packages
- R Studio interface

Data cleaning or wrangling in R

Analyzing your data

Advanced topics in R or data visualization

- Interactive graphs and charts (Shiny)

What we'll cover

1. Intro to data visualization

2. Base R graphics

   BREAK

3. The Grammar of Graphics and ggplot2

   BREAK

4. Additional packages

5. Wrap Up

Data Visualization

Data visualization
is the graphical display of
abstract information
for two purposes:

sense-making and communication

- Stephen Few
Data Visualization for Human Perception

Examine your data

- What type of data have you acquired?
- Do you have all the data you need?
- Does it have all the variables you are interested in?
- Are there any obvious errors in your data?
- Are you missing any data?
- Is your data in the appropriate format?
- Do you have appropriate permissions to use this data?

Other questions to consider

- What question are you trying to answer?
- What is your data telling you?
  - Trends, patterns, outliers, peaks, valleys, etc
- Who is your audience?
  - What is their level of familiarity with this subject?
  - How much time will they have to look at your product?
  - How will they access your product?

Visualization with Base R

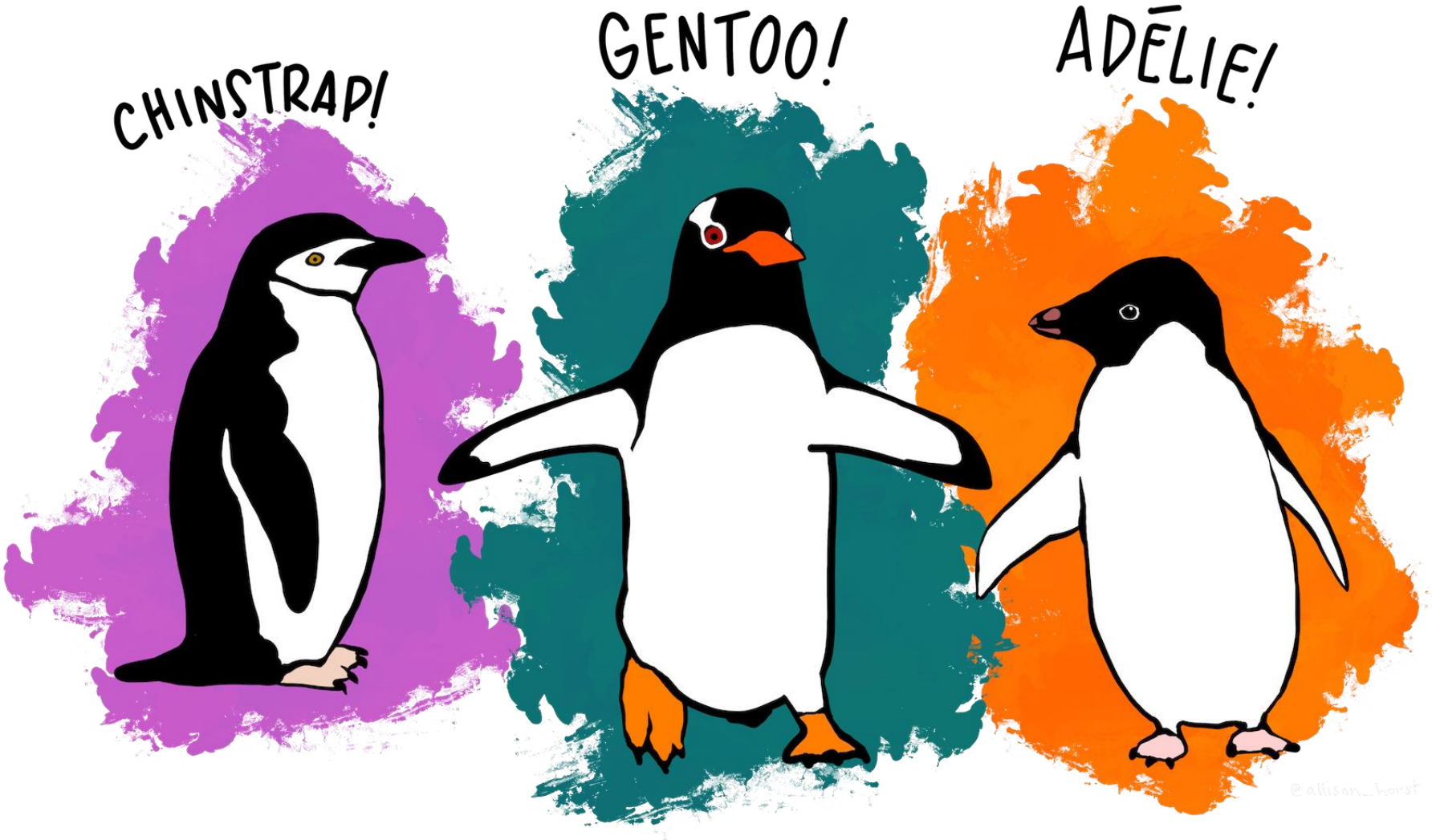
What we'll cover

1. plot()
2. Visualizing 1 variable
3. Visualizing 2 variables
4. Changing graph parameters

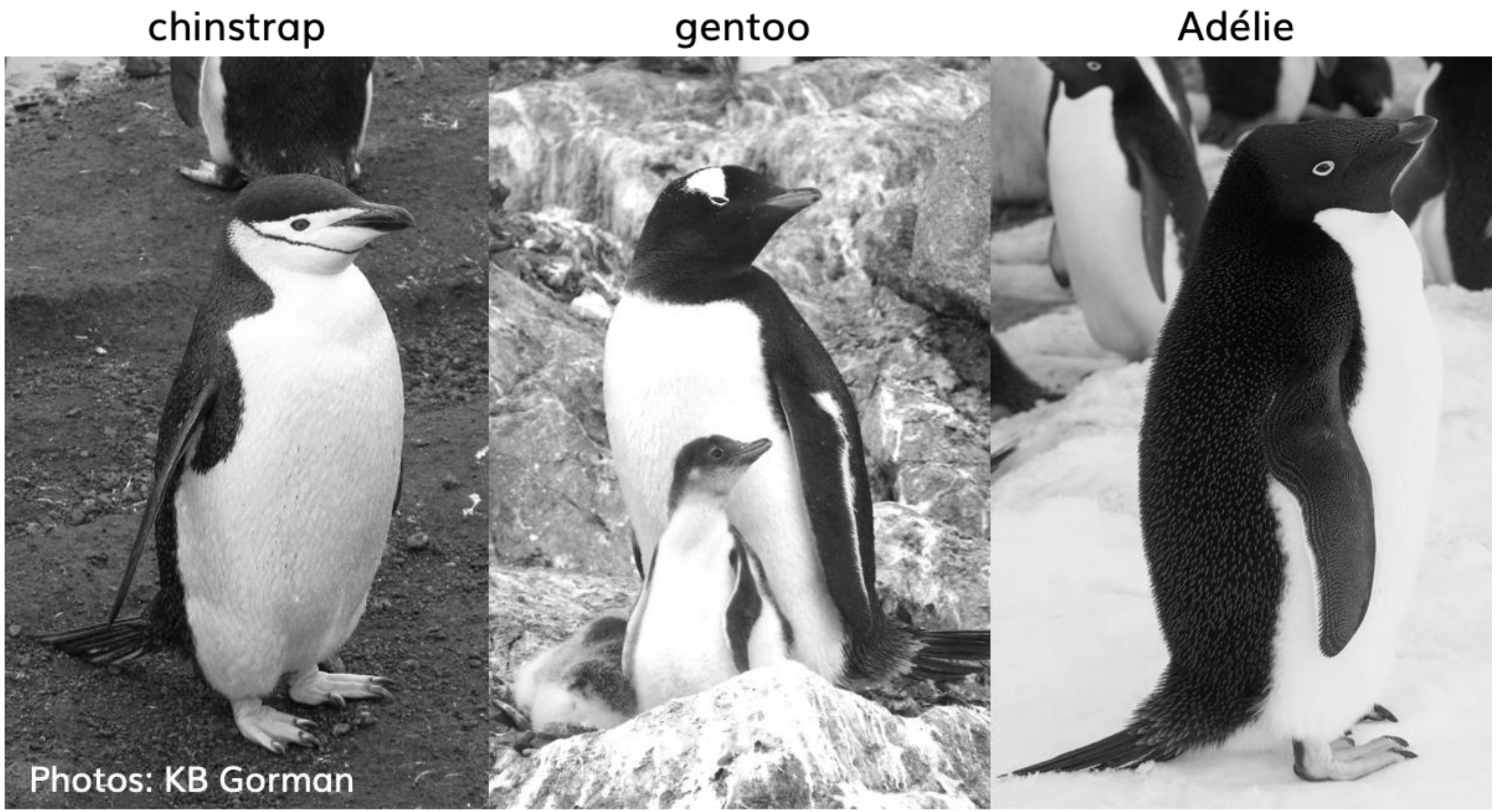
Common Functions in R Graphics

	R graphics
scatterplot	`plot`
bar chart	`barplot`
pie chart	`pie`
histogram	`hist`
box plot	`boxplot`
scatterplot by group	`coplot`

Meet the Penguins!

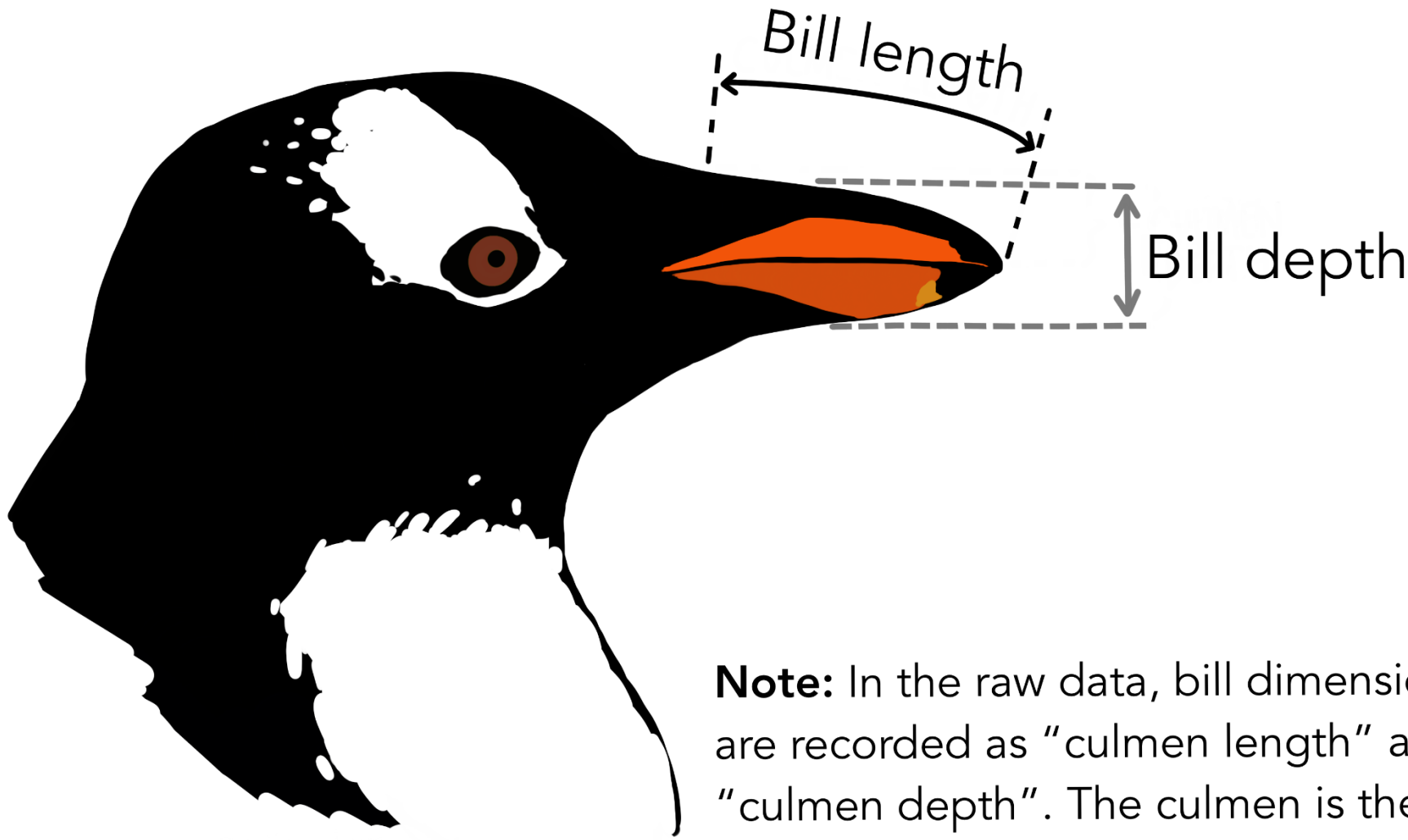


Horst AM, Hill AP, Gorman KB (2020). palmerpenguins: Palmer Archipelago (Antarctica) penguin data. R package version 0.1.0. https://allisonhorst.github.io/palmerpenguins/. doi: 10.5281/zenodo.3960218.



The three penguin species in the *palmerpenguins* R package. Photos: KB Gorman.

Dr. Kristen Gorman in the field with penguins. Photo: S. Sternbach.



Note: In the raw data, bill dimensions are recorded as "culmen length" and "culmen depth". The culmen is the dorsal ridge atop the bill.

Visualization with ggplot2

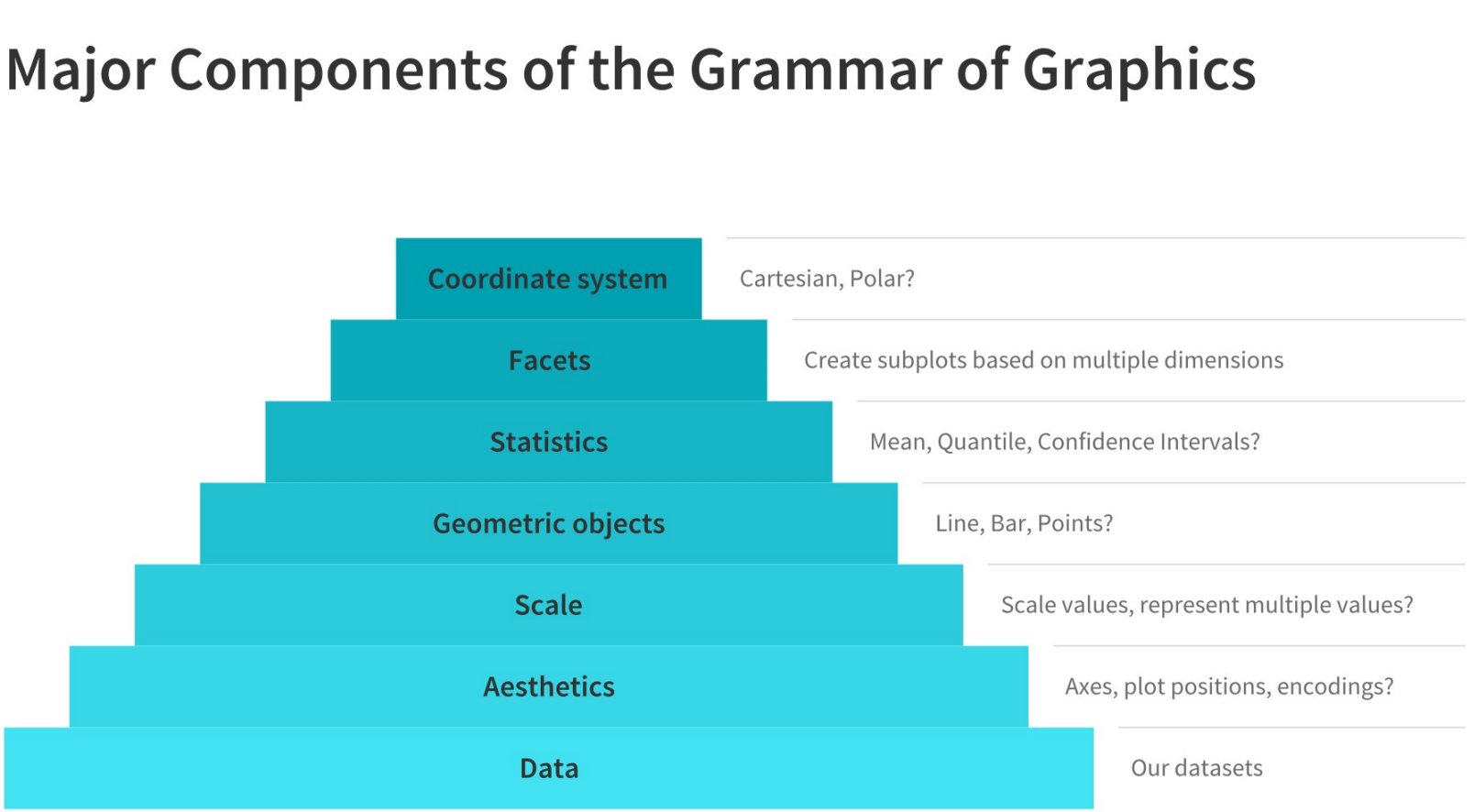
ggplot2

- Popular package for data visualization
- Created by Hadley Wickham in 2005
- Part of the tidyverse package
- Implements the ideas for graphics introduced in "The Grammar of Graphics" by Leland Wilkinson
- Takes care of 'fiddly' details (ie: legends, axes)

https://ggplot2.tidyverse.org/reference/

Grammar of Graphics

Every graph can be described as a combination of independent building blocks:



https://towardsdatascience.com/a-comprehensive-guide-to-the-grammar-of-graphics-for-effective-visualization-of-multi-dimensional-1f92b4ed4149

Grammar of Graphics

Every graph can be described as a combination of independent building blocks:

DATA	a data frame
AES	Aesthetic mapping of variables into visual properties (color, pattern, size, etc)
GEOM	Geometric objects representing the data (bars, lines, points, areas, etc)
COORD	Coordinate system (Cartesian, log, polar, map, etc)
STAT	Statistical transformation – data summaries (mean, standard deviation, etc)
SCALES	Legends, axes, etc
FACETS	Subgrouping capabilities

JOHNS HOPKINS
LIBRARIES
Data Services

Common functions in base R

Type	R graphics
scatterplot	`plot`
bar chart	`barplot`
pie chart	`pie`
histogram	`hist`
box plot	`boxplot`
Scatterplot by group	`coplot`

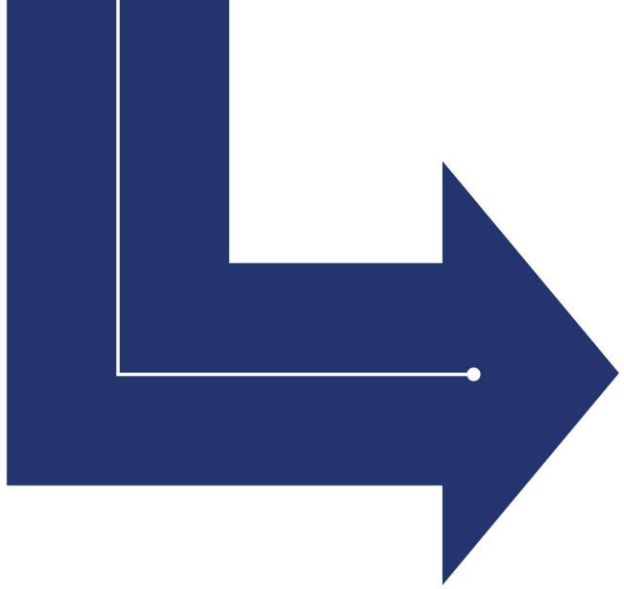
Common functions in ggplot2

Type	R graphics	ggplot2
scatterplot	`plot`	`qplot`
bar chart	`barplot`	`ggplot`
pie chart	`pie`	`ggplot` with `polar_coord` argument
histogram	`hist`	`ggplot` with `geom ="histogram"`
box plot	`boxplot`	`ggplot` with `geom="boxplot"`
Scatterplot by group	`coplot`	`ggplot` with `facets` argument

What we'll cover

1. Plot multiple variables
2. Plot summary statistics
3. Change the aesthetics of a plot
4. Add context to a plot (title, axis labels, etc)
5. Subplots

Additional Packages



Preliminary Data Exploration

- visdat
- DataExplorer

Special Formats

- Lattice (Trellis graphs – multivariate data)
- Leaflet for R (maps)
- Simple features for R (maps)
- RGL (3D visualizations)
- Shiny (interactive)

Other

- dygraphs (Javascript library)
- Plotly (Javascript library)
- Esquisse (plug-in for ggplot2)

Wrap Up

What we learned

1. Looked at basic concepts of data visualization

2. Explored the terminology and syntax for base R and ggplot2

3. Created some basic plots and graphs using base R and ggplot2

What should I use?

- Comparing ggplot2 and R Base Graphics by Nathan Yau
- Why I don't use ggplot2 by Jeff Leek
- Why I use ggplot2 by David Robinson
- There is no ‘right’ way to make plots in R
  - Try out different packages
  - Pick what works for you

Homework

- Practice what we did today
- Review the documentation for base R graphics and ggplot2
- Check out the resources and other tutorials shared after this workshop
- **Practice, practice, practice**

THANK YOU!

Take our survey so we can continue to improve and also expand our offerings:

https://www.surveymonkey.com/r/datavis_r

Data Visualization in R

Instructor: **Peter Lawson**

JOHNS HOPKINS
LIBRARIES
Data Services

Website: dataservices.library.jhu.edu

Email: dataservices@jhu.edu

JHU Data Archive: archive.data.jhu.edu