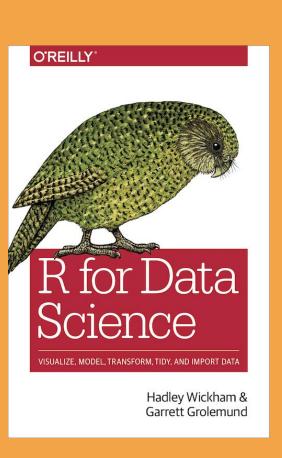
# R for Data Analysis

Salvador Balkus President, Big Data Club



For your reference...

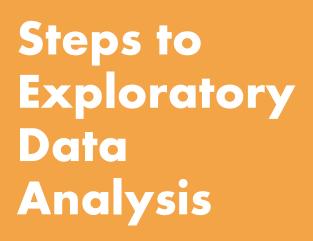


# The most important package

### Tidyverse Cornerstones

Click a tile to access the package documentation and read what it does!







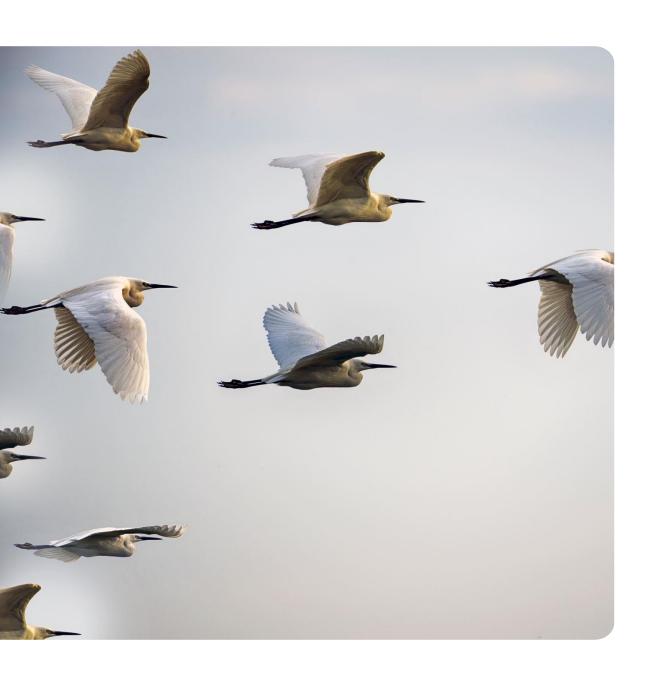
Generate Questions



Find answers by summarizing and visualizing



Refine and generate new questions



### **Our Data:**

**Birds** 

### How do we begin?







Get a sense for the what the variables look like



Make a simple visualization using **ggplot2** 



## Some terminology before proceeding

**Data frame**: a table of data which we want to analyze

- Each row is a variable or attribute
- Each column is an observation
- Also called a "tibble"; foundation of the tidyverse

## Data Visualization with ggplot2

Start with **ggplot(***data***)** - creates a coordinate system

Add one or more layers using **geom** functions

e.g. geom\_bar(...), geom\_histogram(...)

Use **aes(...)** to define the variables visualized

Fancify graph with labels, legends, and themes

# How to make a ggplot

#### **Format:**

```
ggplot(data) + some geometry +
Labels + theme
```

For example, to create a green histogram with a white background...

```
ggplot(data) +
geom_histogram(aes(x = variable),
fill = "green") + xlab("Variable
Name") + theme_light()
```



## Data Transformation: Types of transformations include...



**filter(...)** - choose only certain observations



arrange(...) - sort observations



**select(...)** - choose only certain columns



**mutate(...)** - create new variable from old



**transmute(...)** - edit existing variable

### Split-Apply-Combine:

We can analyze data by group

Split Split into groups using **group\_by(...)** Apply Apply a summary statistic with **summarize(...)** Combine Display the data frame by group

# The Pipe: %>%

- Use to chain commands together
- Output of previous given as input to next

#### Example:

df %>% filter(variable1)
%>% arrange(variable2)

### Dealing with missing values



drop\_na(...)
remove NA observations



replace\_na(...)

replace NA observations with a value

# Thanks for listening! Any questions?