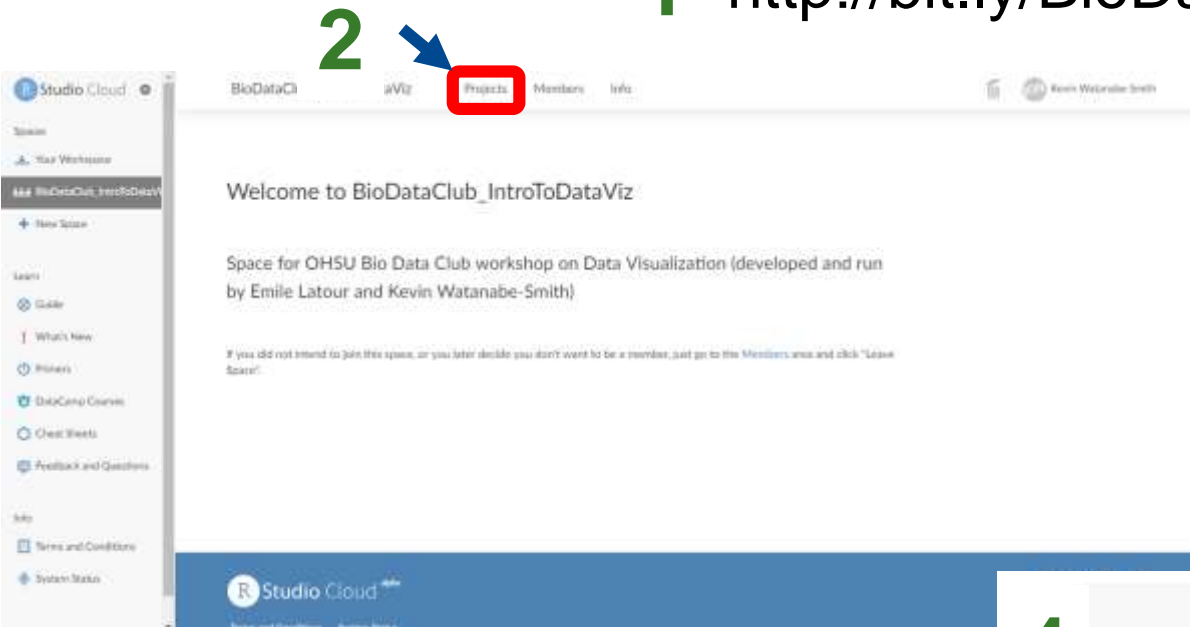
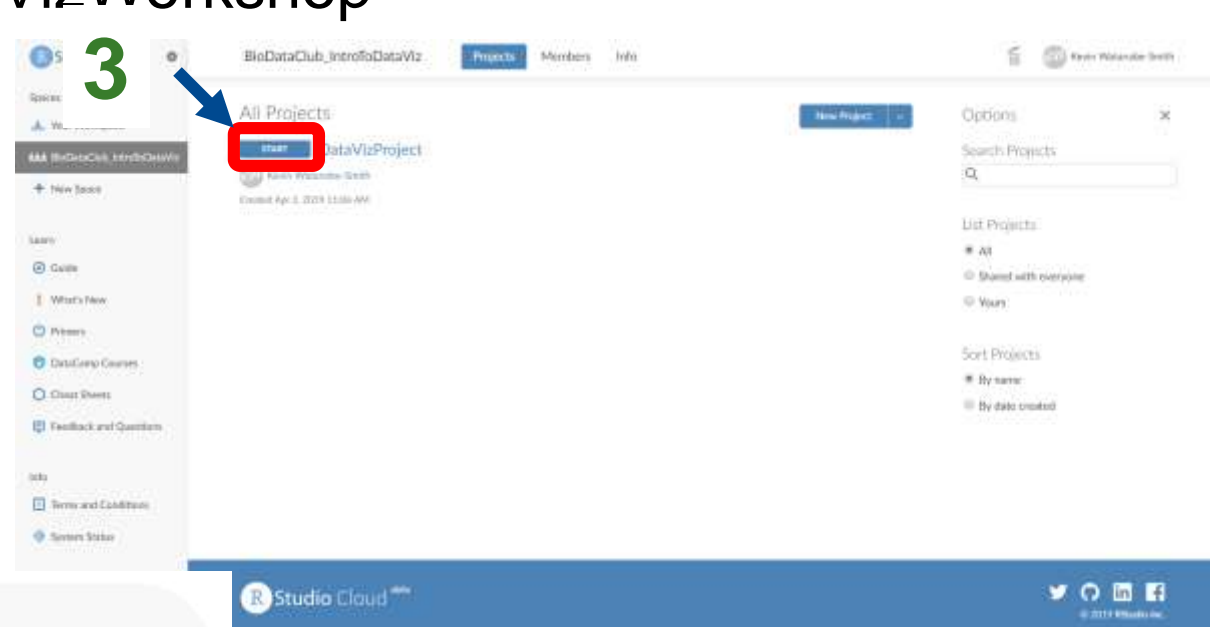


1 <http://bit.ly/BioDataVizWorkshop>

2

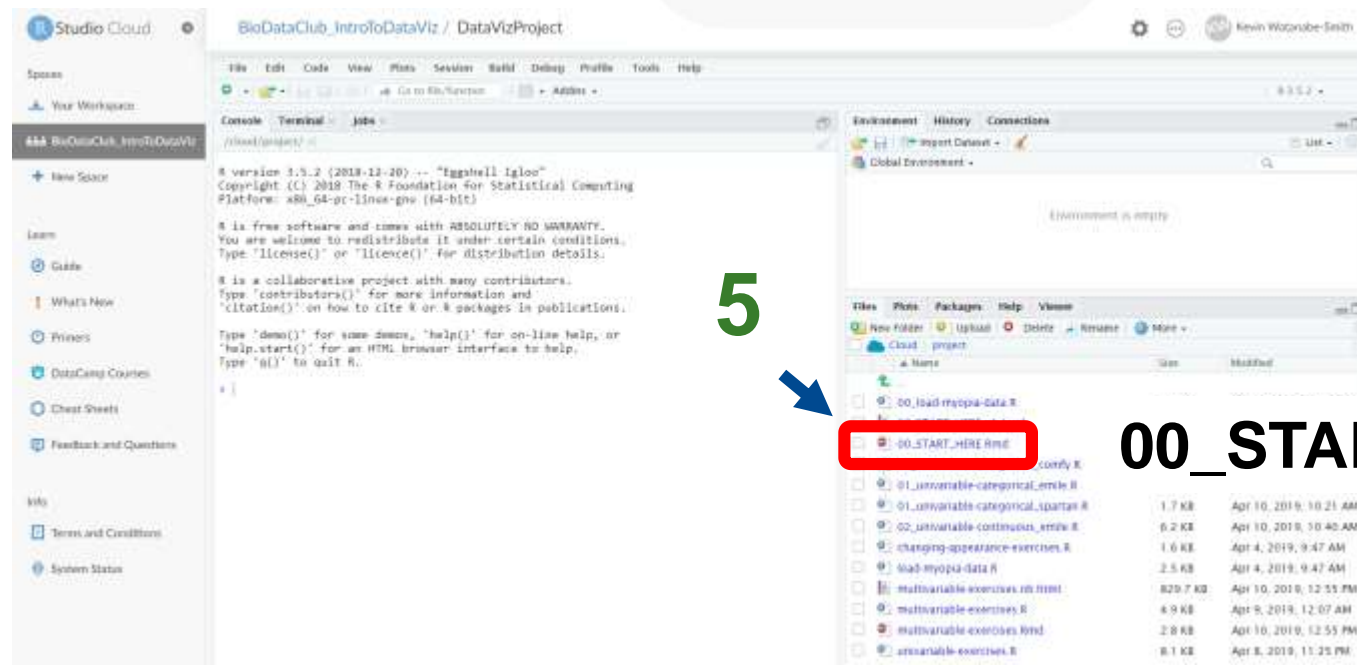


3



4 Deploying Project

5



File Name	Size	Modified
00_load_myopu-data.R	2.5 KB	Apr 4, 2019, 9:47 AM
01_univariable-categorical_emile.R	1.7 KB	Apr 10, 2019, 10:21 AM
01_univariable-categorical_spartan.R	6.2 KB	Apr 10, 2019, 10:40 AM
02_univariable-continuous_emile.R	1.6 KB	Apr 4, 2019, 9:47 AM
02_univariable-continuous_spartan.R	2.5 KB	Apr 4, 2019, 9:47 AM
03_multivariable-exercises.R	829.7 KB	Apr 10, 2019, 12:55 PM
04_multivariable-exercises.R	4.9 KB	Apr 9, 2019, 12:07 AM
05_multivariable-exercises.Rmd	2.8 KB	Apr 10, 2019, 12:55 PM
06_univariable-exercises.R	8.1 KB	Apr 8, 2019, 11:25 PM

00_START_HERE.Rmd

A Practical Introduction to Data Visualization with R and ggplot2

OHSU Biodata club

April 12, 2019

Emile Latour

Kevin Watanabe-Smith

Why code?

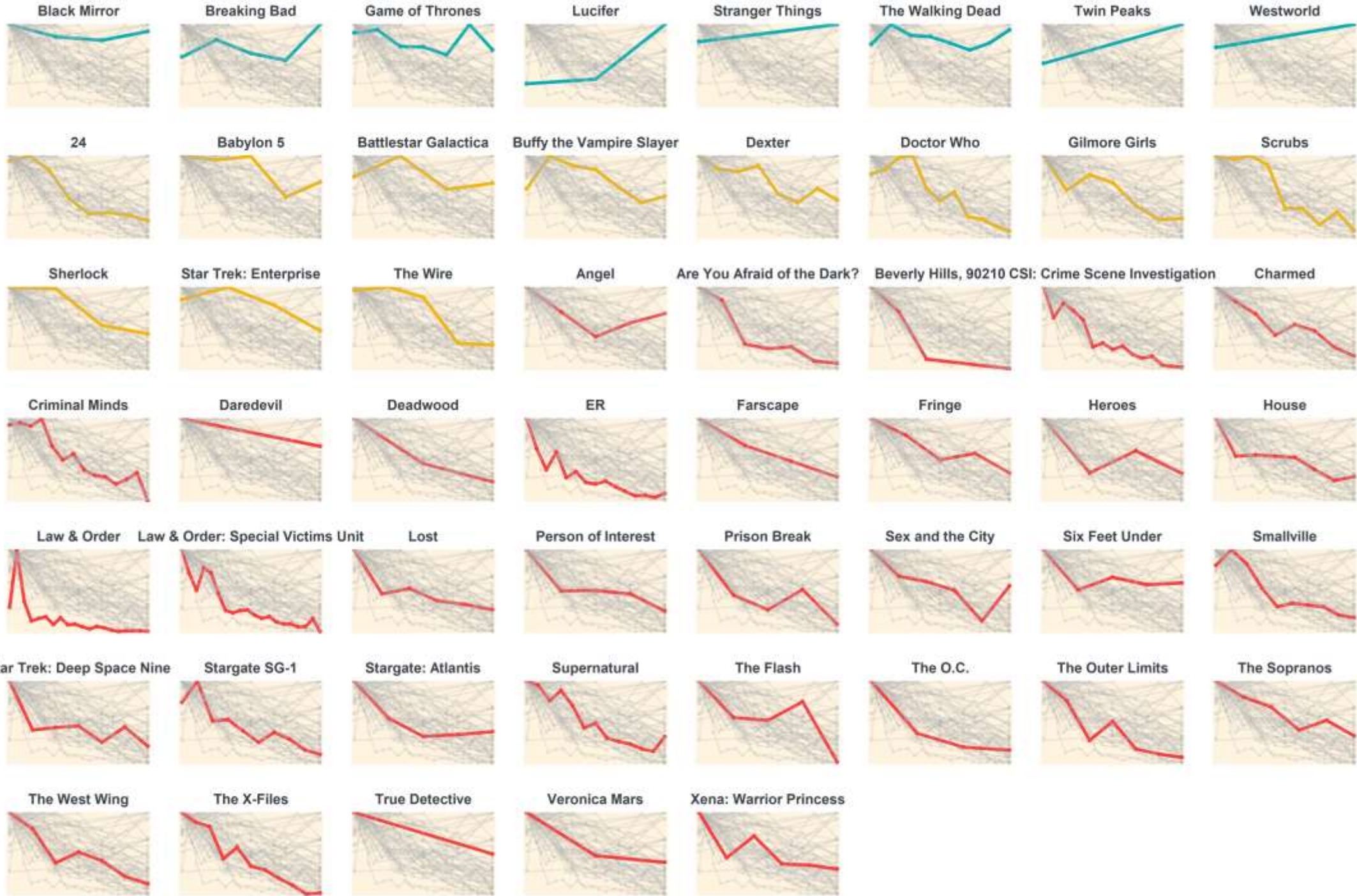
Code has unparalleled expressiveness.

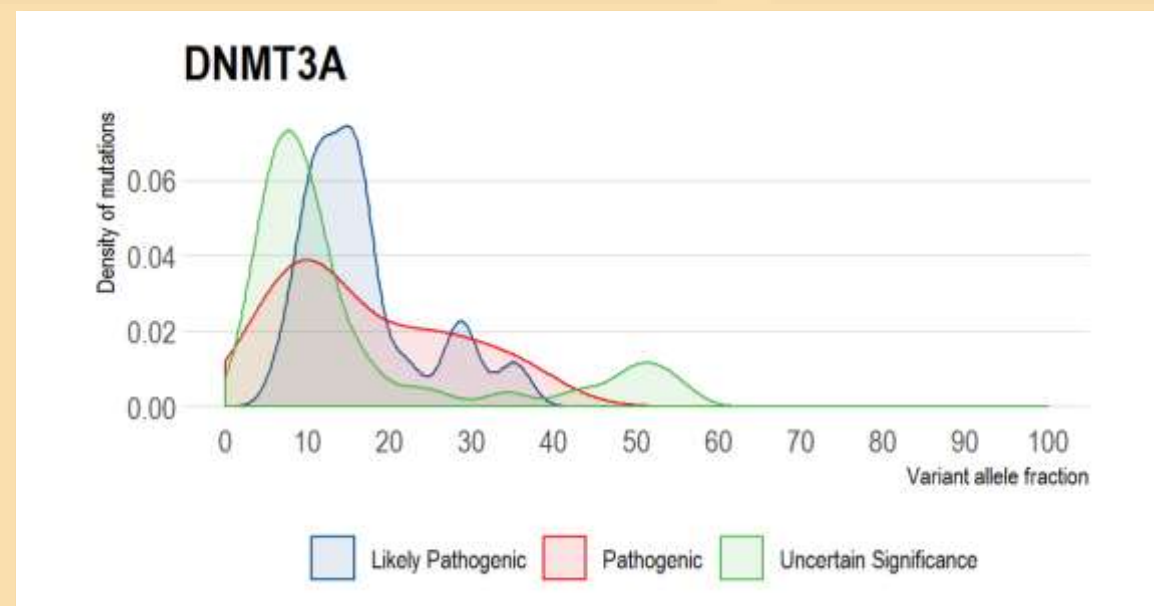
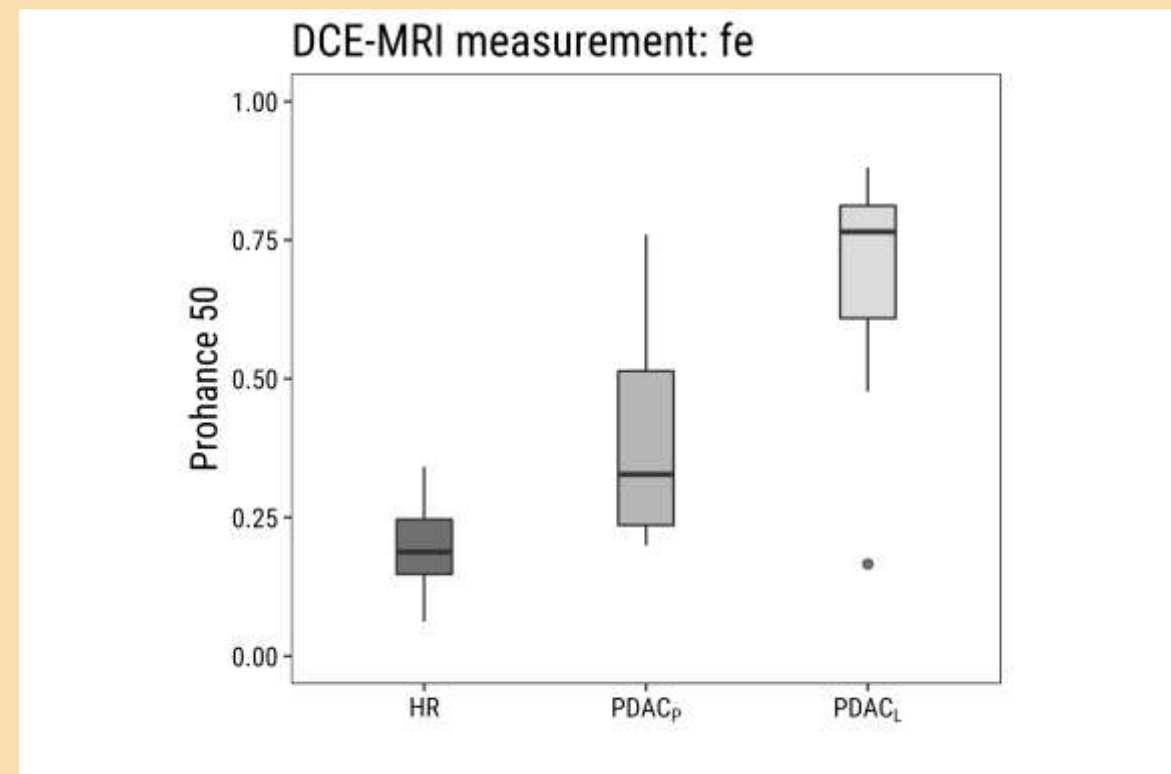
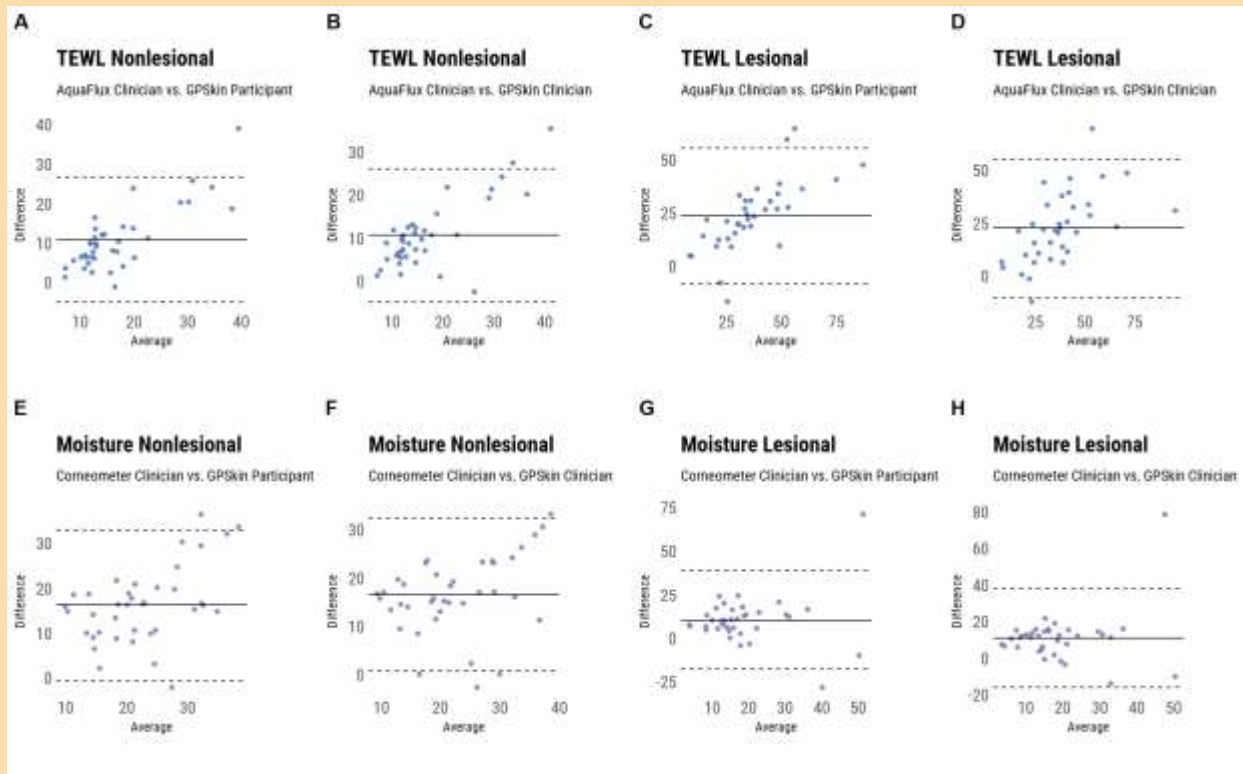
Code is the most general tool we have.
A medium for discovery *must* be general.
Creativity requires composition.

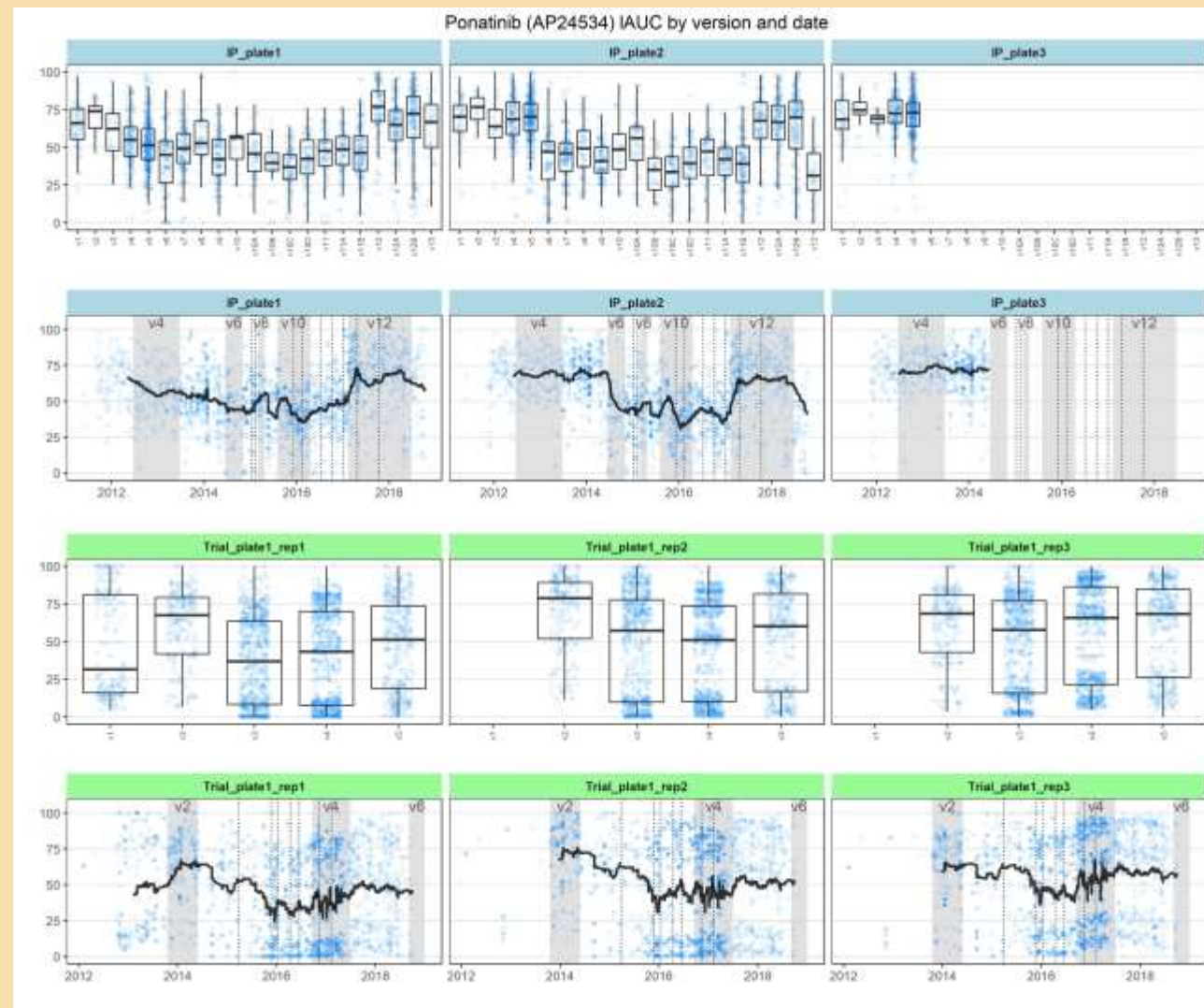
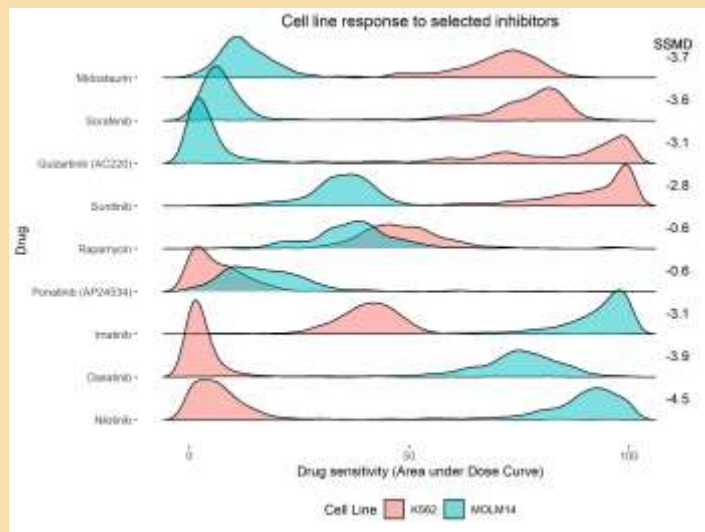
Mike Bostock – CSVConf 2017

Why code?



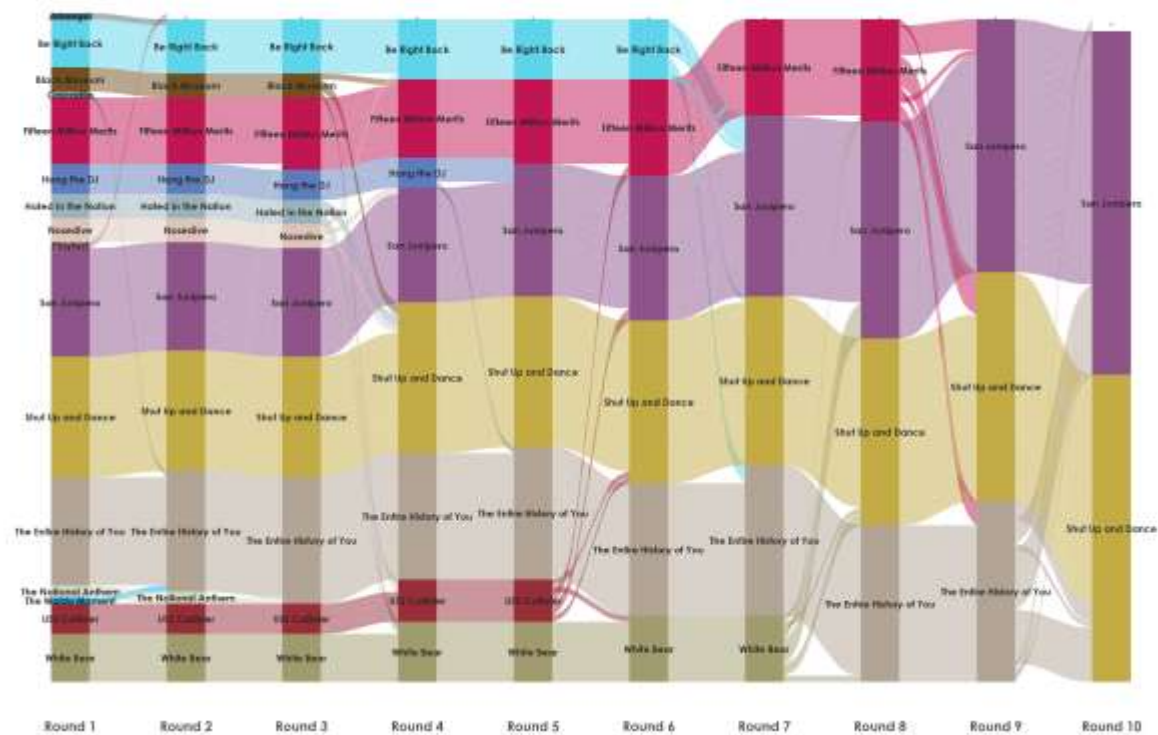






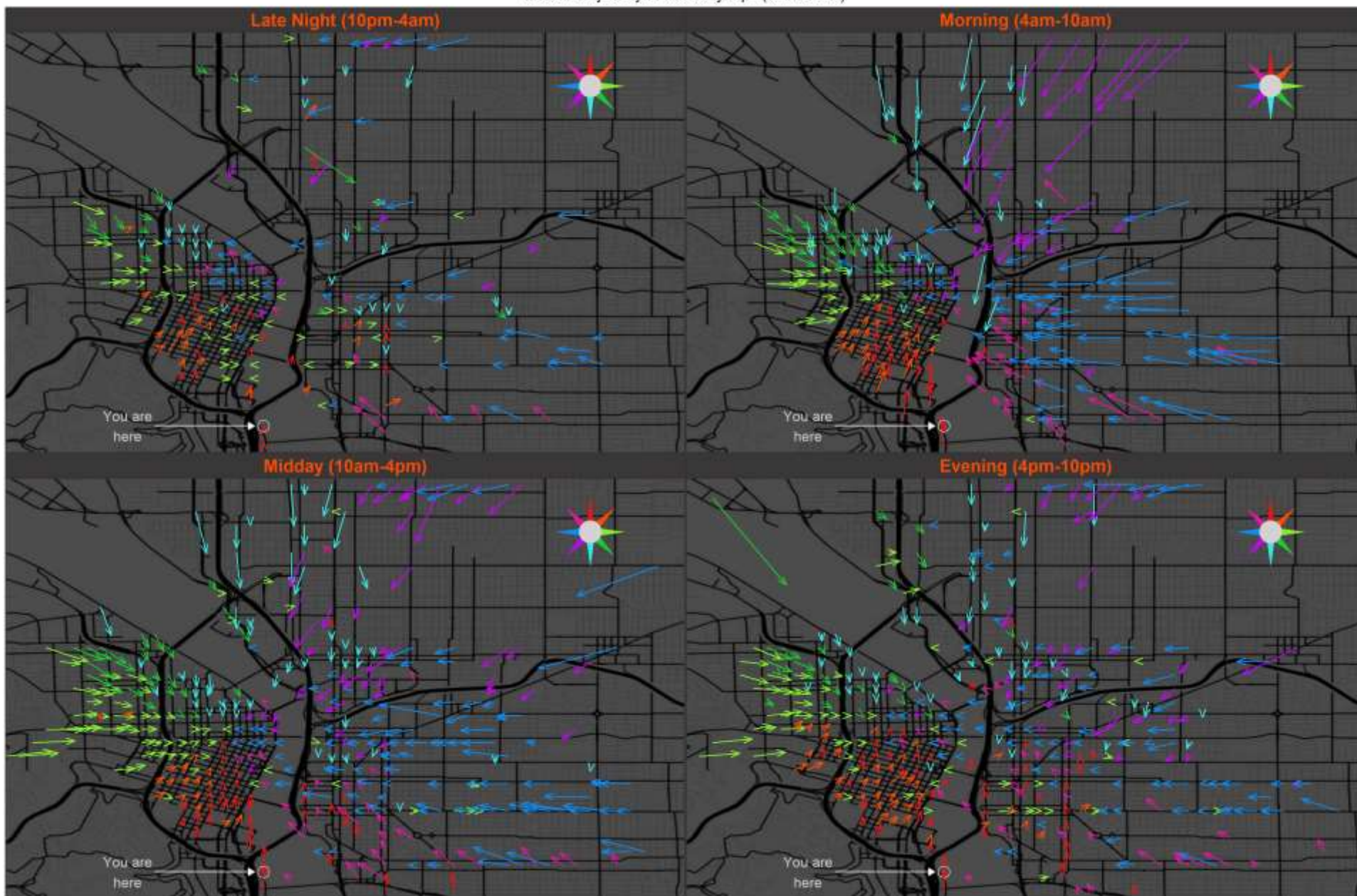
Ranked Choice Voting for Black Mirror Episodes

Non-Jon Hamm edition

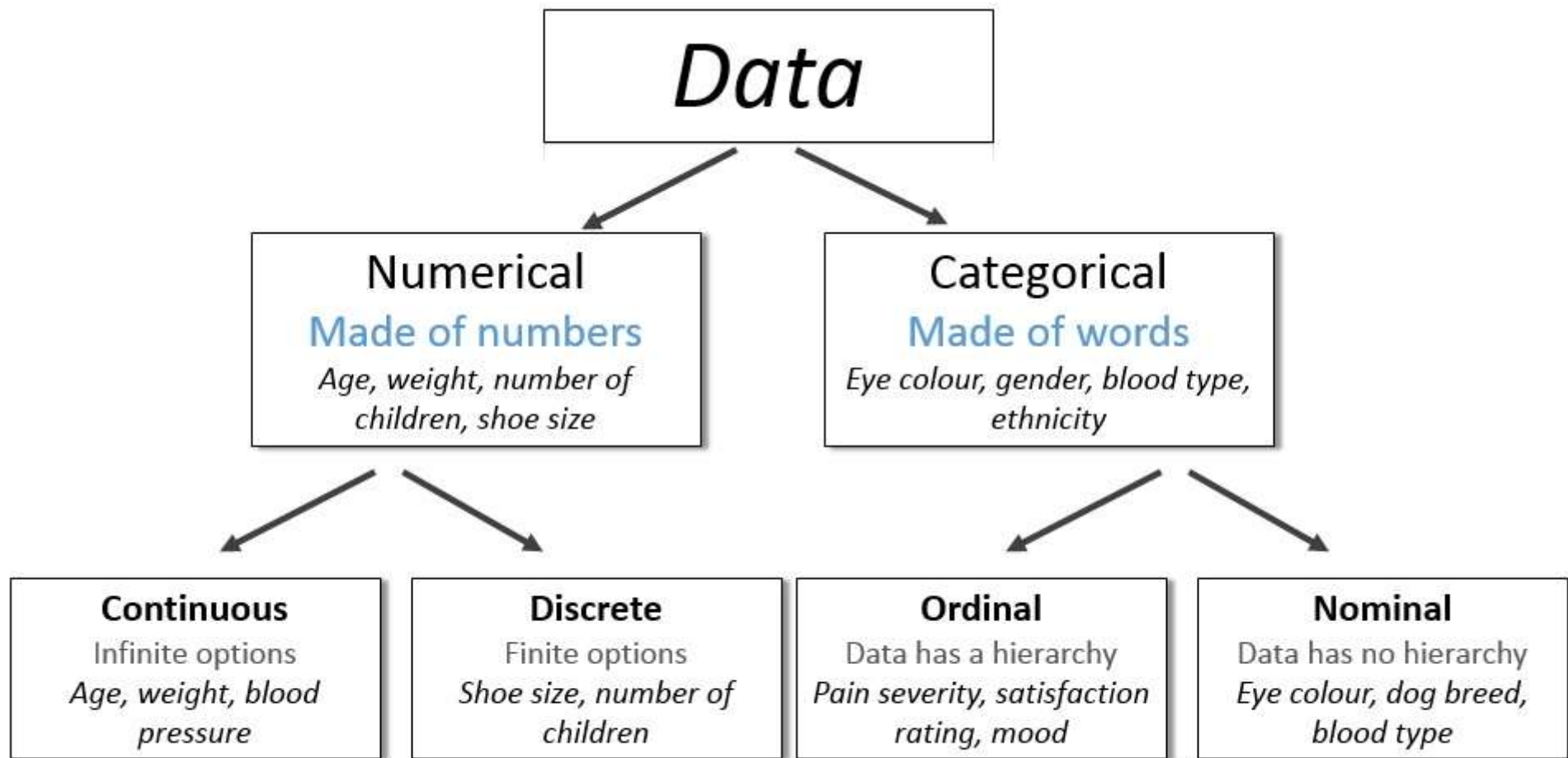


AVERAGE BIKETOWN TRIPS

Median trajectory of weekday trips (2016-2018)



Arrows represent median destination of 10 or more rides starting within a one-block radius (length of rides are 1/4 scale)
R & Rstudio, tidyverse & ggplot2, lubridate, ggmap & Stamen maps, ggrepel, extrafont, BIKETOWN system data
Kevin Watanabe-Smith for CascadiaRConf 2018 - github.com/WatanabeSmith/BIKETOWN_CascadiaR



Univariate plots

Density plots

Histograms & boxplots

Ordered barcharts

Barcharts

Two Variables

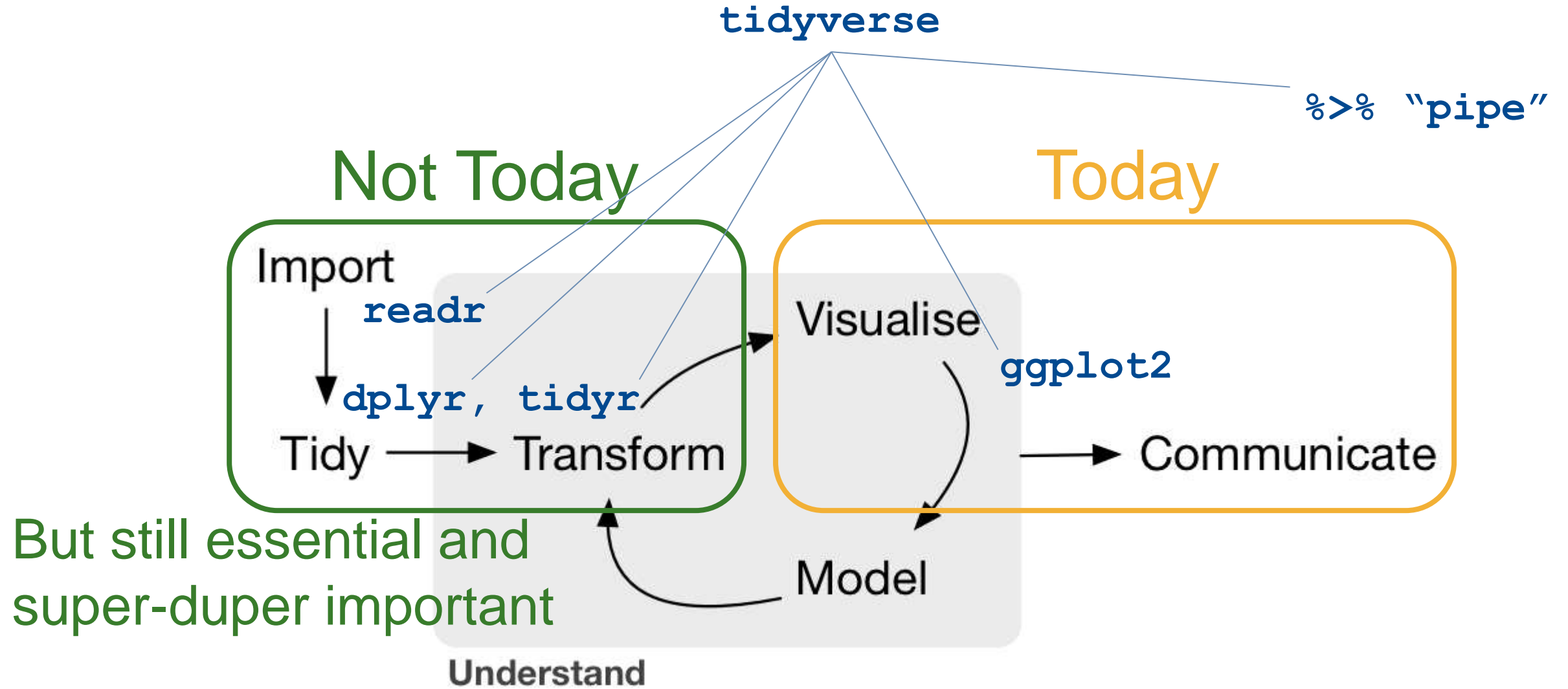
Continuous vs Continuous: Scatterplot Line graph

Continuous vs Categorical: Many boxplots

Categorical vs Categorical: A table Heatmap

More than Two Variables





Alright, now to ggplot

A coordinate system (we'll skip this mostly)

An “**aesthetic**” or how to map some value in your data to the chart space

e.g. x is the first number, y is the second number
x is age, y is income, gender is color

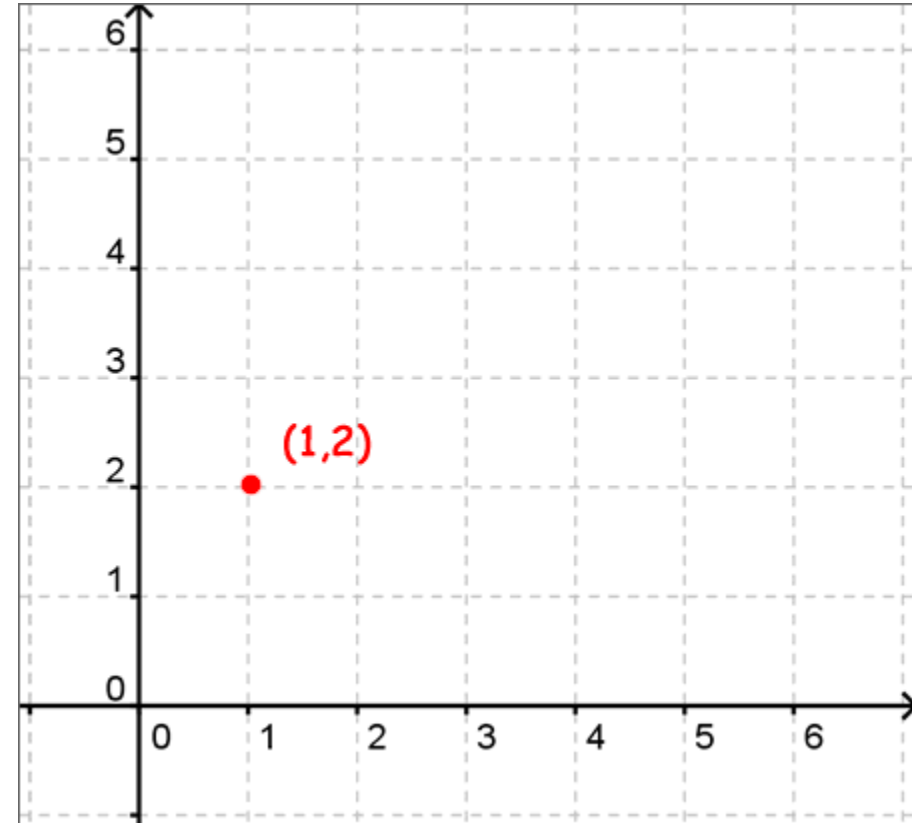
A **geometric object** (geom) or what to plot

e.g. a point, a line, a boxplot

Facets – should the chart be separated into several, smaller charts

Design choices (**theme**) – everything else to make your chart look how you want it

e.g. really boring with lots of tick mark labels, an excess of gridlines, and no text titles




Alright, now to ggplot


plot	ggplot (data = yourdataset,
aesthetic	mapping = aes (x = var_x,
	y = var_y)) +
geom	geom_point ()


ggplot (yourdataset,
aes (var_x,
var_y)) +
geom_point ()

Layer: geoms

A layer combines data, aesthetic mapping, a geom (geometric object), a stat (statistical transformation), and a position adjustment. Typically, you will create layers using a `geom_` function, overriding the default position and stat if needed.


 `geom_abline()` `geom_hline()` `geom_vline()` Reference lines: horizontal, vertical, and diagonal


 `geom_bar()` `geom_col()` `stat_count()` Bar charts


 `geom_bin2d()` `stat_bin_2d()` Heatmap of 2d bin counts

 `geom_blank()` Draw nothing

 `geom_boxplot()` `stat_boxplot()` A box and whiskers plot (in the style of Tukey)

 `geom_contour()` `stat_contour()` 2d contours of a 3d surface

 `geom_count()` `stat_sum()` Count overlapping points

 `geom_density()` `stat_density()` Smoothed density estimates

`geom_density_2d()` `stat_density_2d()` Contours of a 2d density estimate

 `geom_dotplot()` Dot plot

`geom_errorbarh()` Horizontal error bars

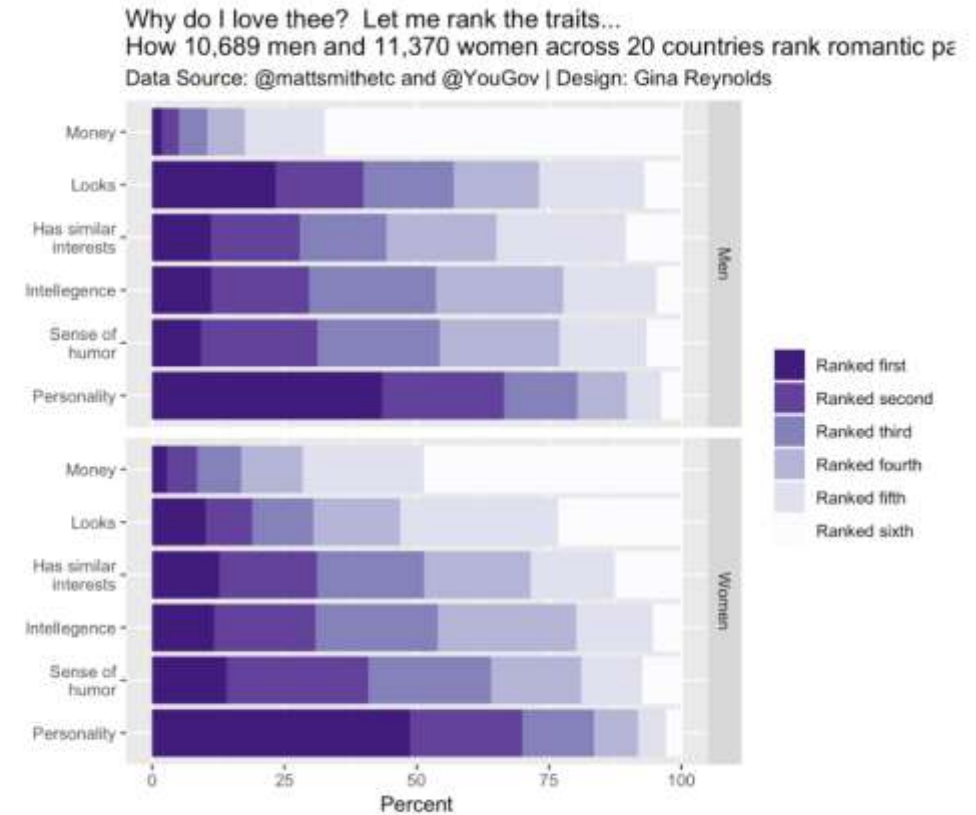
 `geom_hex()` `stat_bin_hex()` Hexagonal heatmap of 2d bin counts

Very useful resources

```
ggplot(data = world) +  
  aes(x = Question_short_wrap, y = Percent) +  
  aes(fill = 'Rank (text)') +  
  facet_grid(Gender ~ .) +  
  geom_col() +  
  coord_flip() +  
  scale_fill_manual(values = colorRampPalette(RColorBrewer::brewer.pal(9, "Purples"))(6))  
  labs(fill = "") +  
  xlab("") +  
  labs(title = "Why do I love thee? Let me rank the traits...", subtitle = "How 10,689 men and 11,370 women across 20 countries rank romantic partner traits")  
  labs(subtitle = "Data Source: @mattsmithetc and @YouGov | Design: Gina Reynolds")
```

evamaerey.github.io/ggplot_flipbook/ggplot_flipbook_xaringan.html

<http://sape.inf.usi.ch/quick-reference/ggplot2/colour>



Cheatsheet



<https://ggplot2.tidyverse.org/index.html>

Very useful books

How to actually learn any new programming concept



Essential

Changing Stuff and
Seeing What Happens

O RLY?

@ThePracticalDev

Solutions that might fix the problem without breaking anything



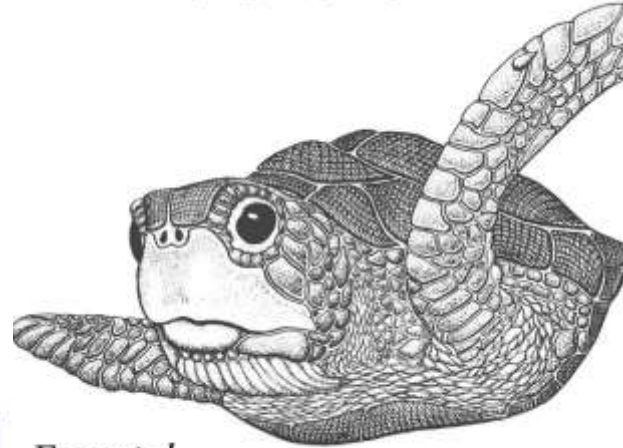
Essential

Hoping This
Works

O RLY?

@ThePracticalDev

How you get anything done ever



Essential

Remembering
What to Google

O RLY?

@ThePracticalDev

Cutting corners to meet arbitrary management deadlines



Essential

Copying and Pasting
from Stack Overflow

O'REILLY®

The Practical Developer
@ThePracticalDev

Very useful books

<https://kieranhealy.org/publications/dataviz/>

How to actually learn any new programming concept Solutions that



Essential

Changing Stuff and
Seeing What Happens

O RLY?

@ThePracticalDev

Essential

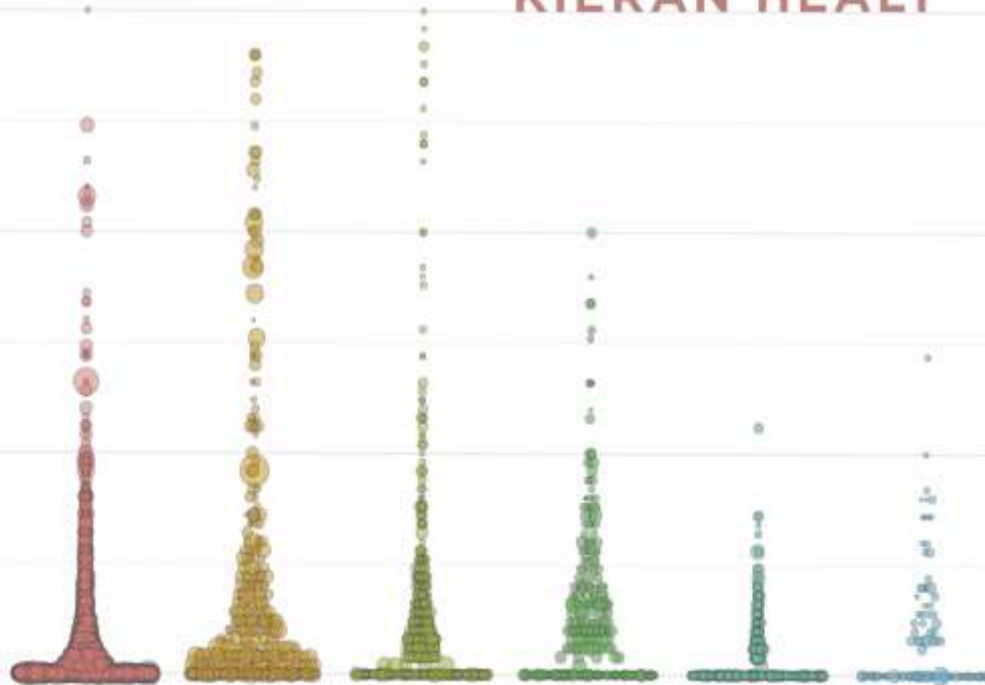
H
W

O RLY?

DATA VISUALIZATION

A PRACTICAL INTRODUCTION

KIERAN HEALY



Cutting corners to meet arbitrary management deadlines



Essential

Copying and Pasting
from Stack Overflow

O'REILLY®

The Practical Developer
@ThePracticalDev

<http://bit.ly/BioDataVizWorkshop>

The left screenshot shows the RStudio Cloud interface for the workspace 'BioDataClub_IntroToDataViz'. The 'Projects' tab is highlighted with a red box and a blue arrow. The right screenshot shows the 'All Projects' view for the same workspace, with the 'DataVizProject' highlighted by a red box and a blue arrow.

Deploying Project

The screenshot shows the RStudio Cloud interface for the workspace 'BioDataClub_IntroToDataViz / DataVizProject'. The 'Files' pane on the right shows a list of files, with '00_START_HERE.Rmd' highlighted by a red box and a blue arrow.

00_START_HERE.Rmd

Click around, change the size of each pane, make it yours
This is on a webpage, it can't break your computer and I promise you won't break Rstudio cloud (probably)

Studio Cloud

Spaces

Your Workspace

BioDataClub_IntroToDataViz

+ New Space

Learn

Guide

What's New

Primers

DataCamp Courses

Cheat Sheets

Feedback and Questions

Info

Terms and Conditions

System Status

BioDataClub_IntroToDataViz / DataVizProject

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function

Addins

R 3.5.2

Environment History Connections

Import Dataset

Global Environment

Files Plots Packages Help Viewer

New Folder Upload Delete Rename

Cloud project

File	Size
00_load-myopia-data.R	2.7 KB
00_Start_Here.Rmd	694.3 KB
00_START_HERE.Rmd	534 B
01_univariable-categorical_com...	3.1 KB
01_univariable-categorical_emil...	2.9 KB
01_univariable-categorical_spar...	1.7 KB
02_univariable-continuous_emil...	8.2 KB
changing-appearance-exercises.R	1.6 KB
load-myopia-data.R	2.5 KB
multivariable-exercises.html	829.7 KB
multivariable-exercises.Rmd	2.8 KB
univariable-exercises.R	8.1 KB

Source

This is where you type stuff (code, text) that you want to keep
Also where code already written will be loaded in

```
1 ---
2 title: "00_START_HERE"
3 output: html_notebook
4 ---
5
6 Welcome to the BioData Club Introduction to Data Visualization workshop! This is a R Notebook file, which allows you
7 to alternate between typing text (like here) and blocks of R code that you can run (see below)
8
9 # This is a R Notebook file, which allows you to alternate between typing text (like here) and blocks of R code that
10 library(tidyverse)
11 head(ChickWeight)
12
13
14 ggplot(ChickWeight,
15       aes(x = Time,
16           y = weight,
17           by = Chick)) +
18   geom_line() +
19   facet_wrap(~ Chick)
20
21 # Chunk 1
```

Console

This is where you type code, but nothing here will be saved
Good for checking something quickly

Environment

This shows all of your currently stored variables

Everything else

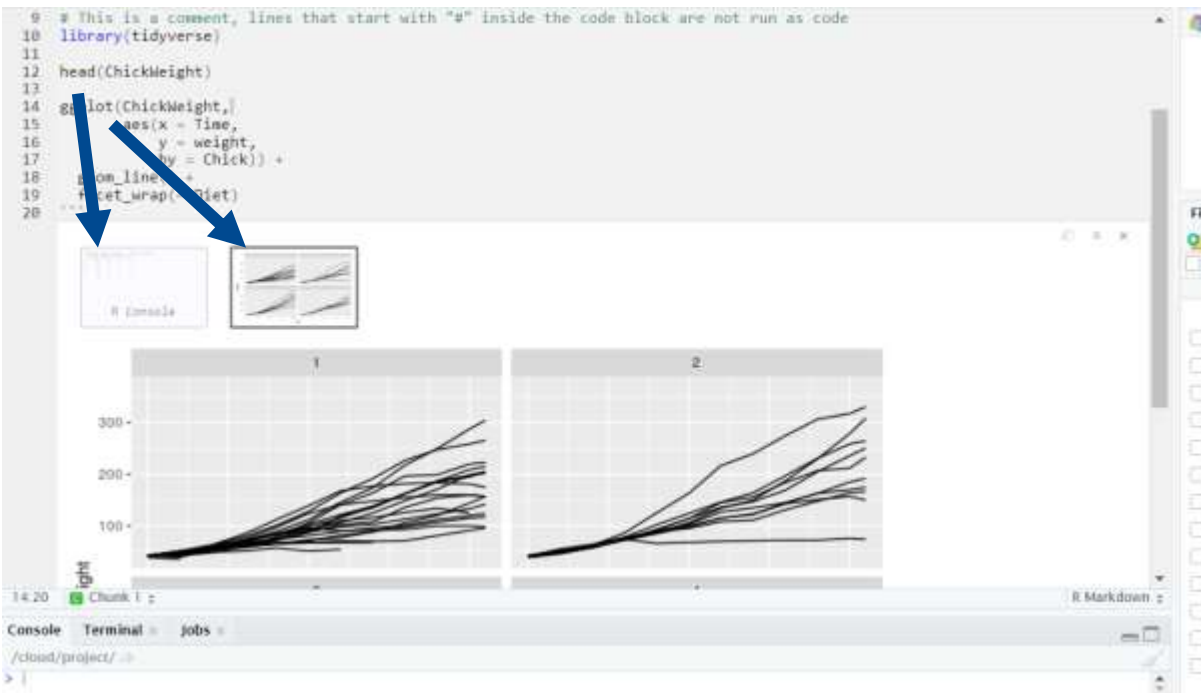
Files in your workspace

Plots you have made

Help and documentation pages

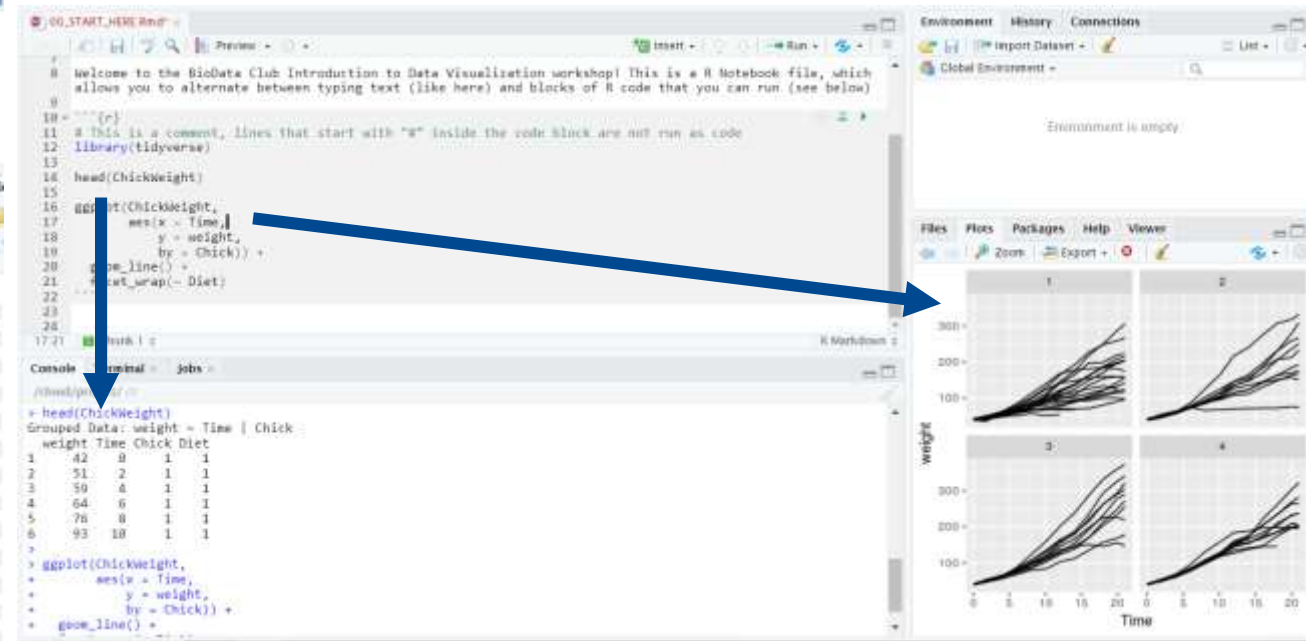


Inline output

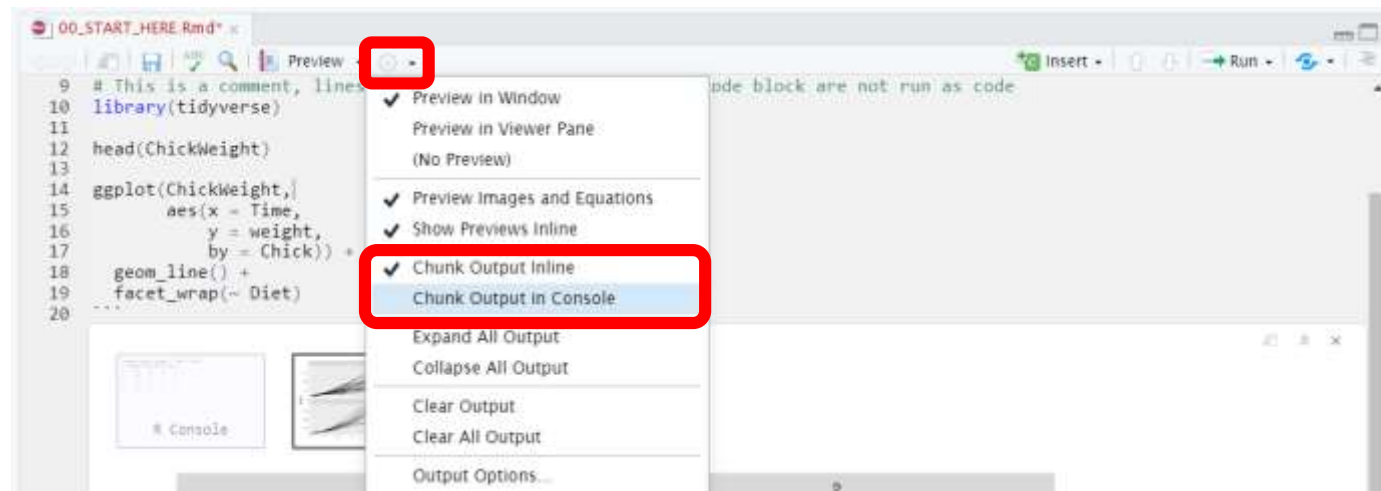


All of your code and output together in one window

Console output



Output text goes to console, charts to Plots pane



Run all the code in the document

Run just the
line my
cursor is on

Run just this code chunk

The screenshot shows an R Markdown editor with two open files: `00_START_HERE.Rmd*` and `multivariable-exercises.Rmd`. The active file contains the following code:

```
1 ---  
2 title: "00_START_HERE"  
3 output: html_notebook  
4 editor_options:  
5   chunk_output_type: console  
6 ---  
7  
8  
9 Welcome to the BioData Club Introduction to Data Visualization work  
allows you to alternate between typing text (like here) and blocks  
10  
11  
12 ```{r}  
13  
14 # This is a comment, lines that start with "#" inside the code block are not run as code  
15  
16 library(tidyverse)  
17  
18 head(ChickWeight)  
19  
20 ggplot(ChickWeight,  
21       aes(x = Time,  
22           y = weight,  
23           by = Chick)) +  
24   geom_line() +  
25   facet_wrap(~ Diet)  
26 ```
```

A context menu is open, showing the following options and shortcuts:

- Run Selected Line(s) `Ctrl+Enter`
- Run Current Chunk `Ctrl+Shift+Enter`
- Run Next Chunk `Ctrl+Alt+N`
- Run Setup Chunk
- Run All Chunks Above `Ctrl+Alt+P`
- Run All Chunks Below
- Run All `Ctrl+Alt+R`

Arrows indicate the following actions:

- A blue arrow points from the "Run all the code in the document" box to the "Run All" option in the menu.
- A green arrow points from the "Run just the line my cursor is on" box to the "Run Selected Line(s)" option in the menu.
- A blue arrow points from the "Run just this code chunk" box to the "Run Current Chunk" option in the menu.
- A green arrow points from the "Run every chunk above this one" box to the "Run All Chunks Above" option in the menu.

The status bar at the bottom shows "10:1 (Top Level) ↕" and "R Markdown ↕".

00_START_HERE.Rmd* x multivariable-exercises.Rmd x

7
8
9 Welcome to the BioData Club Introduction to Data Visualization workshop:
allows you to alternate between typing text (like here) and blocks of R code

10
11
12 ````{r}`
13
14 # This is a comment, lines that start with "#" inside the code block are
15
16 `library(tidyverse)`
17
18 `head(ChickWeight)`
19
20 `ggplot(ChickWeight,`
21 `aes(x = Time,`
22 `y = weight,`
23 `by = Chick)) +`
24 `geom_line() +`
25 `facet_wrap(~ Diet)`
26 `````
27
28
29 ````{r}`
30
31
32

Insert

- R
- Bash
- D3
- Python
- Rcpp
- SQL
- Stan

Insert a new R chunk

Make a new code chunk (for R code)
(or press Ctrl + Alt + i)

30:1 Chunk 2 R Markdown

How to draw an owl

1.



1. Draw some circles

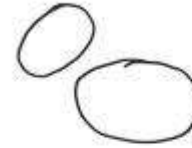
2.



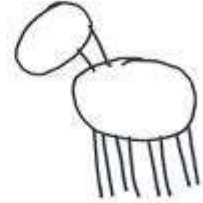
2. Draw the rest of the fucking owl

HOW TO: DRAW A HORSE

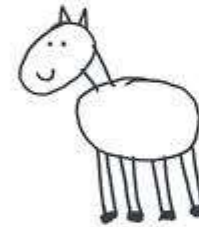
BY VAN OKTOP



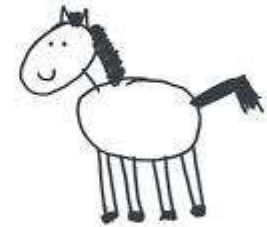
① DRAW 2 CIRCLES



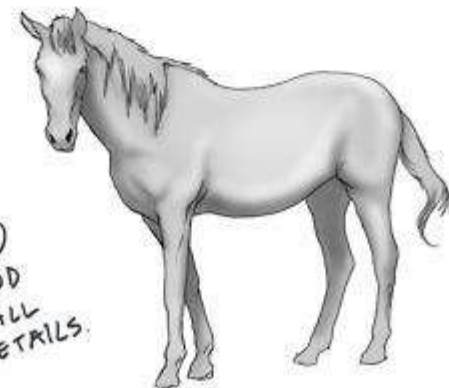
② DRAW THE LEGS



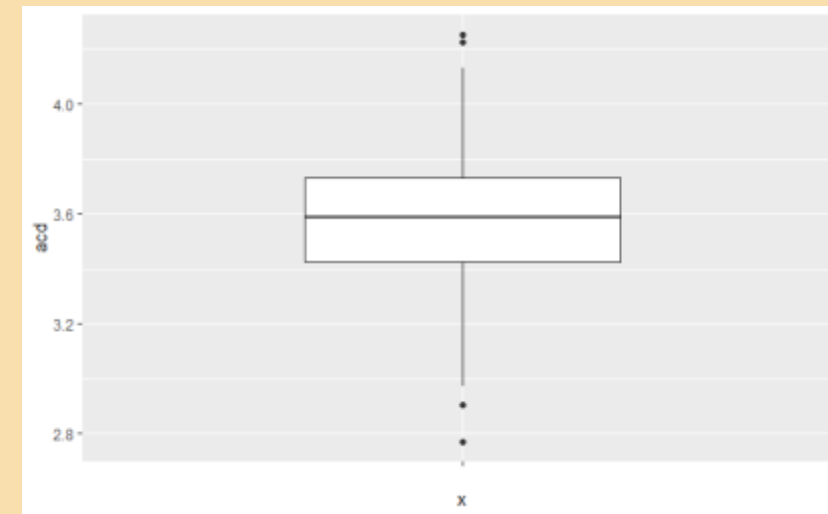
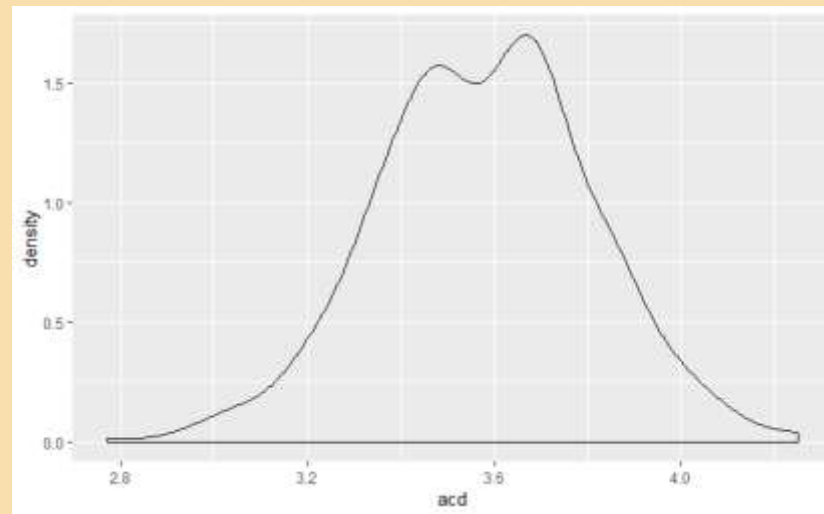
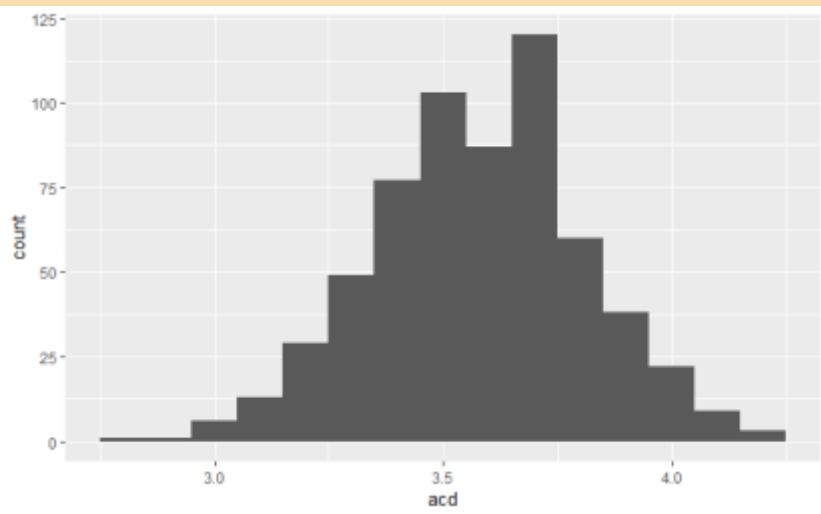
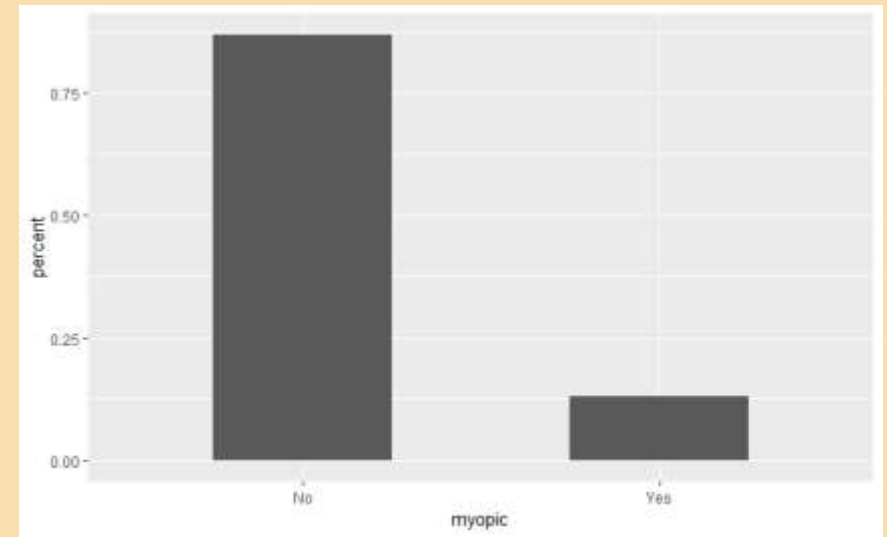
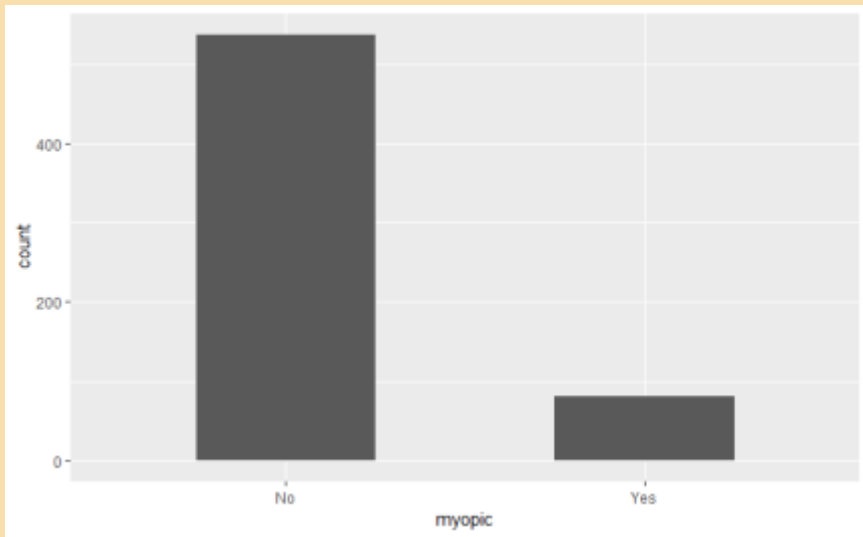
③ DRAW THE FACE

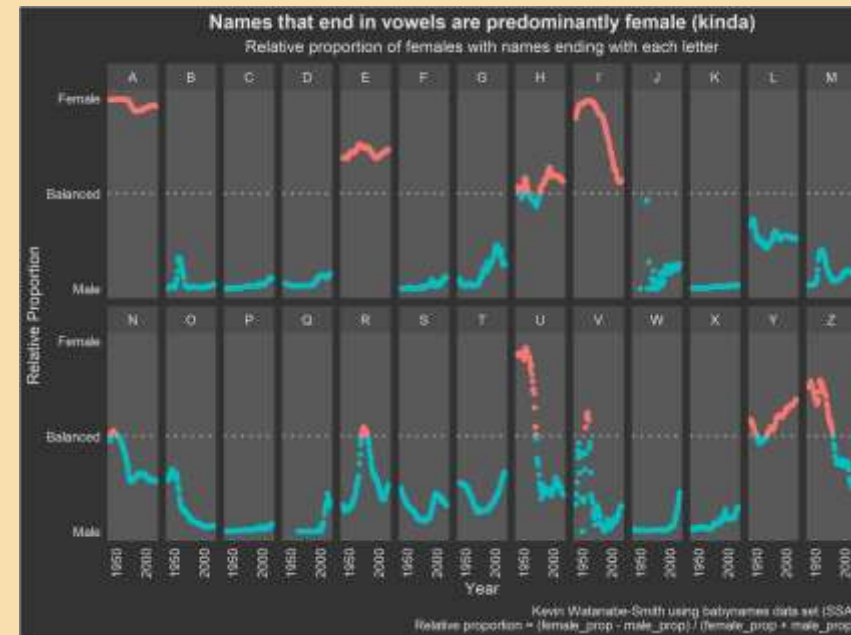
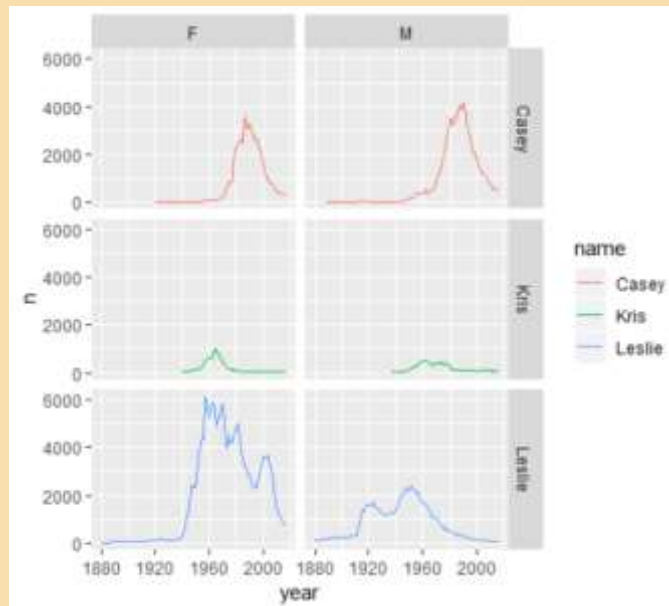
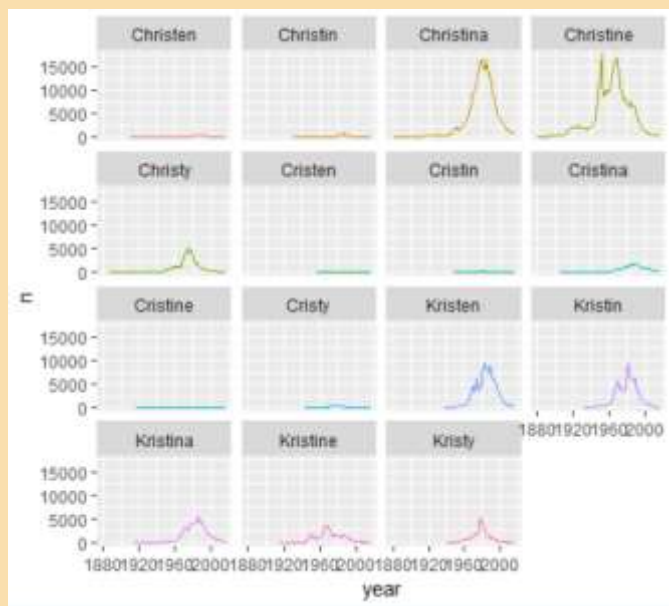
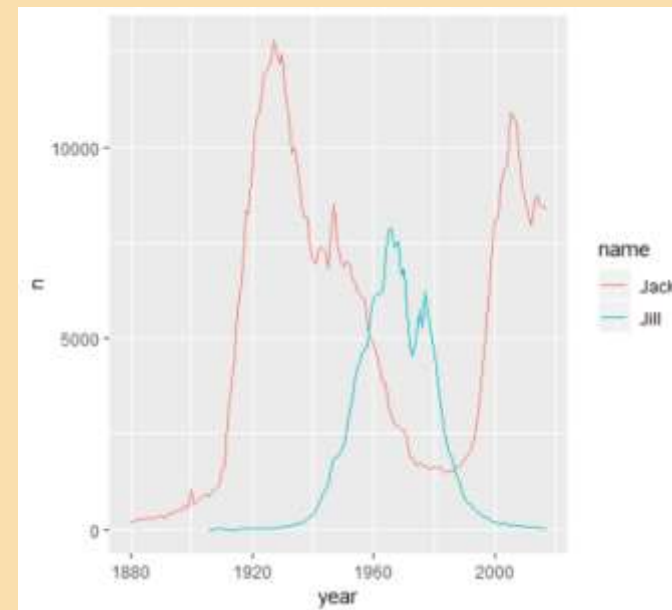
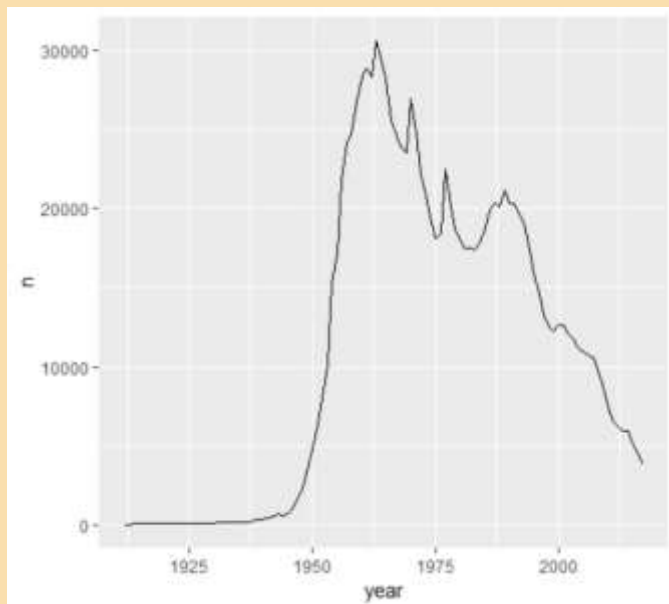


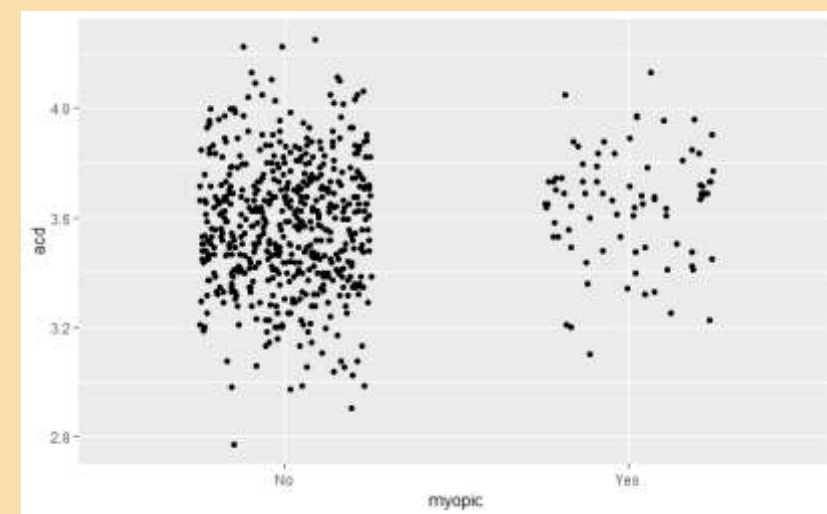
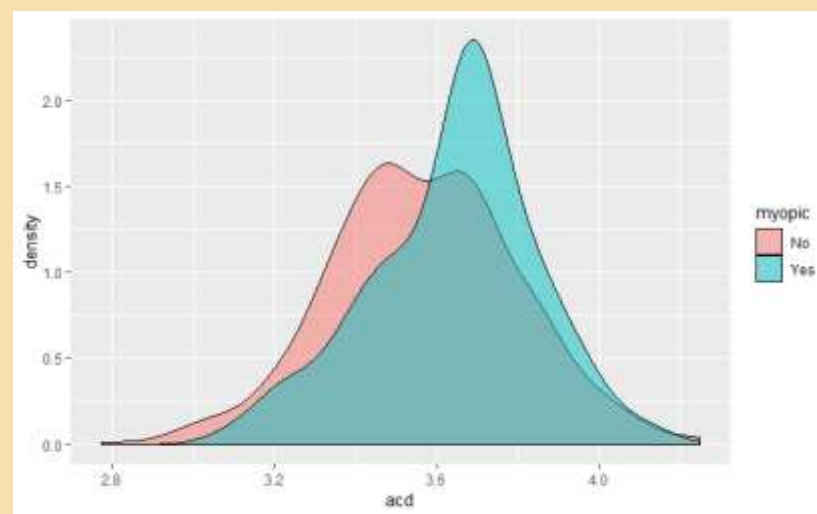
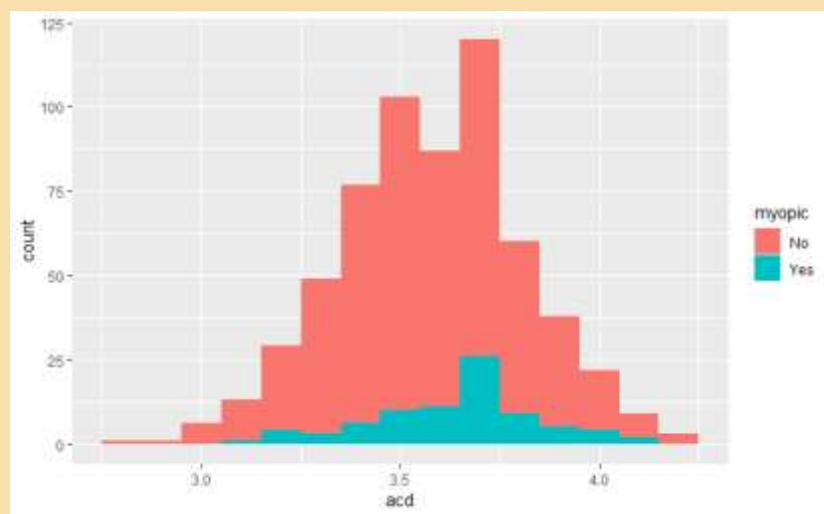
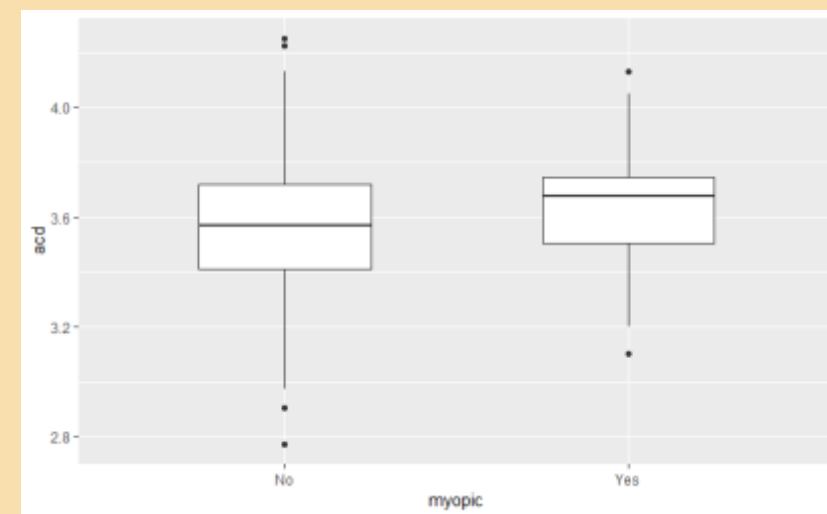
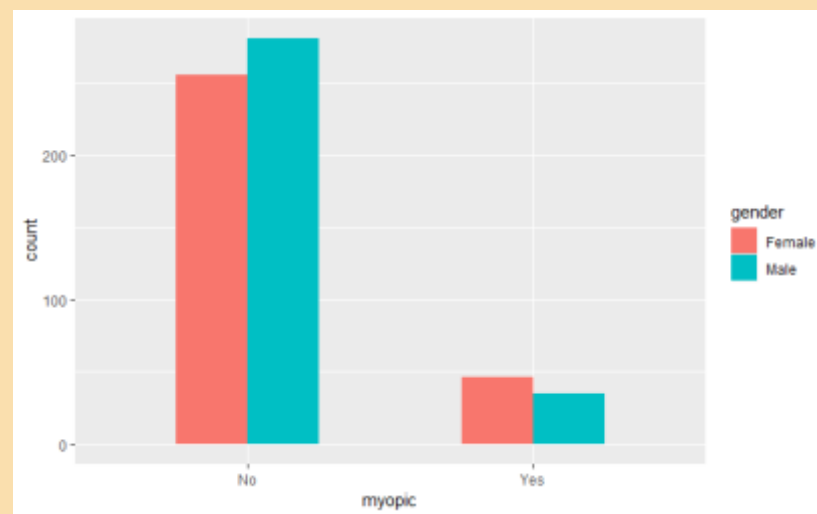
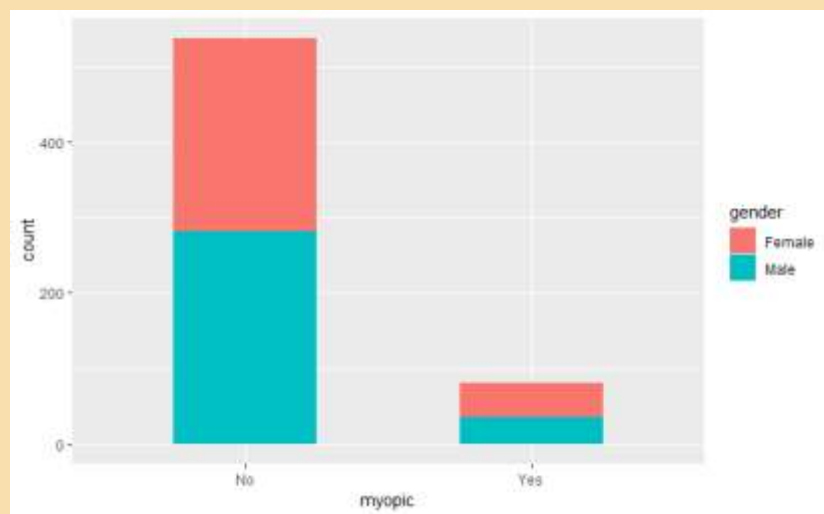
④ DRAW THE HAIR



⑤
ADD
SMALL
DETAILS.

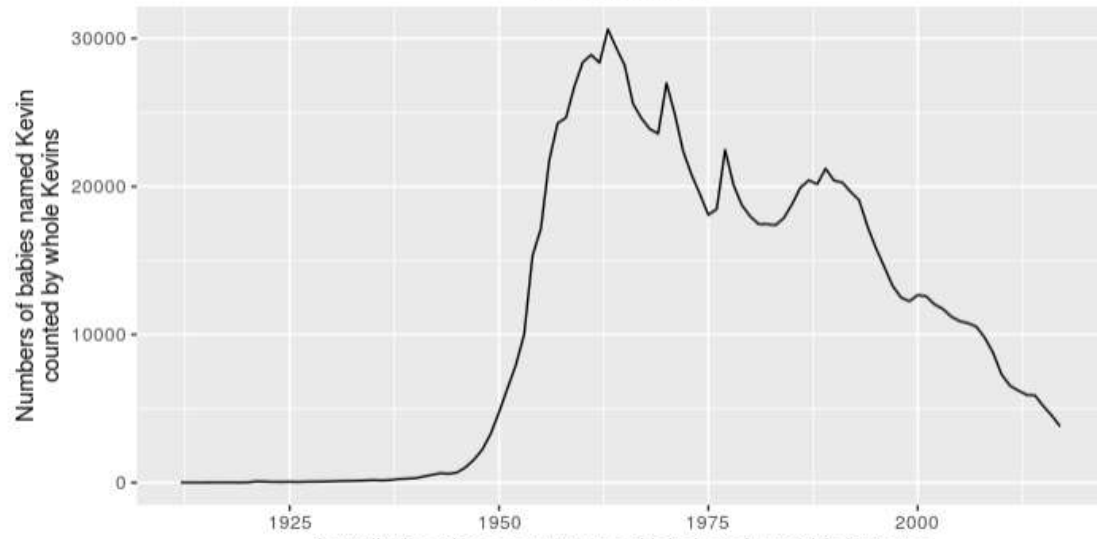






Kevin used to be cool

Not so much now

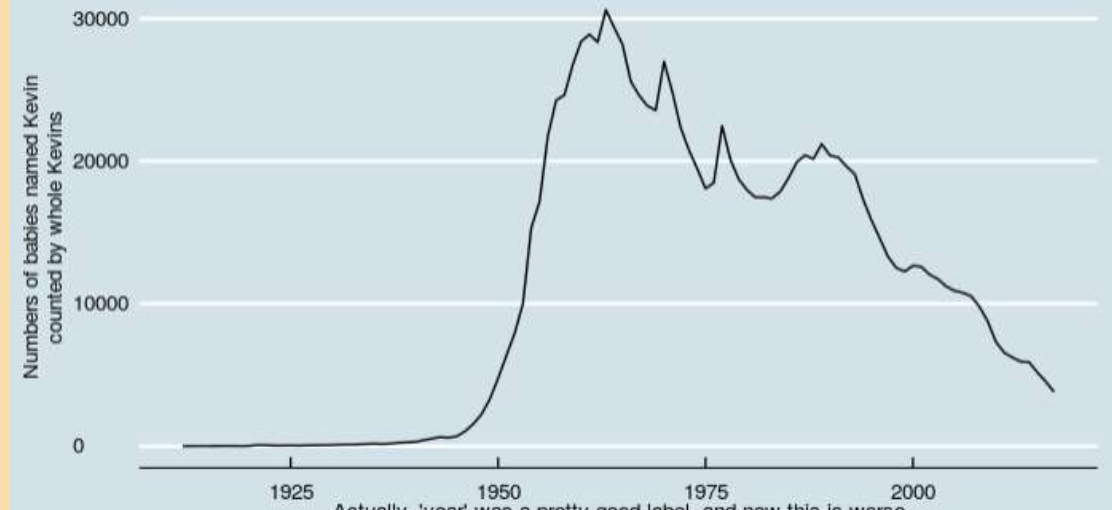


Actually, 'year' was a pretty good label, and now this is worse

by Kevin

Kevin used to be cool

Not so much now

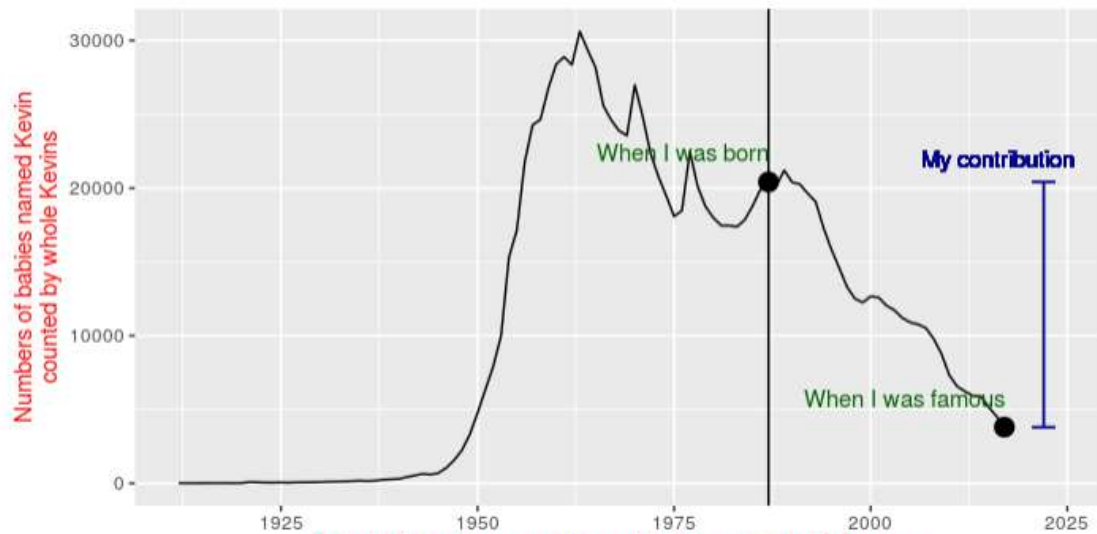


Actually, 'year' was a pretty good label, and now this is worse

by Kevin

Kevin used to be cool

Not so much now



Actually, 'year' was a pretty good label, and now this is worse

by Kevin

