# Developing Data Products Notes

# Coursera Course by John Hopkins University

INSTRUCTORS: Dr. Brian Caffo, Dr. Jeff Leek, Dr. Roger D. Peng

# Contents

Intro	3
Github Link for Lectures	. 3
Course Book	. 3
Instructor's Note	. 3
Shiny, GoogleVis, and Plotly	3
Shiny Part 1	. 3
Shiny Overview (1.1)	. 3
What is Shiny?	. 3
HTML, CSS, and Javascript Tutorials	
A Shiny Project	
Shiny Code Demos (1.2)	
Set-Up	
ui.R	
$\operatorname{server.R}$	
Shiny HTML Tags (1.3)	
Shiny Apps with $I/O$ (1.4)	
Shiny Apps with Plots $(1.5)$	
Shiny Part 2	
Reactivity (2.1)	
reactive Example (2.2)	
Delayed Reactivity (2.3)	
Tabs (2.4)	
Using your own HTML (2.5)	
Interactive Graphics (2.6)	
Shiny Gadgets	
Overview (1.1)	
Shiny Gadgets (1.2)	
Gadgets with Interactive Graphics (1.3)	
GoogleVis	
GoogleVis (1.1)	
Motion Chart Example	
Charts in googleVis (1.2)	
Plots on Maps Example	
Combing Multiple Plots Together	

	Viewing the HTML code	 												12
	Things You Can do with Google Vis	 						 						12
	For More Info	 						 						13
	Plotly													
	Plotly Intro (1.1)													13
	Basic Scatter Plot (1.2)													13
	Adding Qualatative Variables (1.3)													14
	Discreet Color													15
	Continuous Color													16
	Altering Size													16
	3D Scatter Plot													17
														18
	Line Graphs (1.4)													
	Standard Line Graph Example													18
	Multi Line Graph - Stock Example .													19
	Other Plots (1.5)													20
	Histogram													20
	Boxplot													21
	Heatmap													22
	3D Surface													23
	Choropleth Maps $(1.6) \dots \dots \dots$	 												24
	Displaying ggplots with Plotly (1.7)	 												25
	Posting to Plotly (1.8)	 												26
	Quiz 1	 												27
R	Markdown and Leaflet													<b>2</b> 9
	R Markdown	 												
	Overview $(1.1)$	 												29
	Set Up - Presentaion $(1.2)$	 												29
	Editing a Presentaion $(1.3 - 1.5)$	 												29
	Publishing to Github Pages (1.6)	 						 						29
	Sharing R Markdown Documents	 						 						30
	Leaflet	 						 						30
	Leaflet (1.1)	 												30
	Leaflet $(1.2)$													30
	Leaflet $(1.3)$													
	Leaflet (1.4)													30
	Leaflet (1.5)													30
	Leaflet (1.6)													30
	Quiz 2													30
	Course Project 1													30
	Course i roject i	 •	• •	• •	•	•	 •	 •	•	 •	•	•	•	30
В	Building R Packages													31
	R Packages													31
	R Packages Overview													31
	R Packages (1.1)													31
	R Packages (1.2)													31
	Building R Packages Demo													31
	R Classes and Methods													
		 												O.L

	R Classes and Methods (1.1)	3
	R Classes and Methods (1.2)	3.
	uiz 3	31
	ourse Project 2	3.
Sv	·I	31
	wirl	31
	Swirl (1.1)	31
	Swirl (1.2)	3.
	Swirl (1.3)	3.
	ourse Project 3	31

# Intro

Github Link for Lectures

Developing Data Products' lectures on GitHub

Course Book

The book for this course is available on leanpub

#### Instructor's Note

"This course is about building tools for improving the data analysis process, making data driven decisions, or for other infrastructure that supports other data products...

To preview some of the topics we'll be covering. With the R Markdown package you can create websites, PDFs, presentations, and even e-books from a single file in R in a way that you're very comfortable and familiar with. We'll teach the latest features of the Shiny package which you can use to create interactive web applications in R. We'll talk about interactive graphics using Plotly and Leaflet which allows you to create beautiful maps that you can share online. We'll do similar things with the package GoogleViz which allows you to create maps and interactive graphics and tables. We've also added how to use swirl and swirlify to design courses in R so that you can share your knowledge...

- Brian Caffo and the Data Science Track Team"

# Shiny, GoogleVis, and Plotly

Shiny Part 1

Shiny Overview (1.1)

#### What is Shiny?

- Shiny is a web development framework in R, meaning one only needs to know R to use it.
  - Helps one get around a possible lack of resources or knowledge in JavaScipt, HTML, etc.
- Shiny needs a server to run on

- One can use their own, RStudio's limited free hosting service, or something like Amazon AWS. RStudio also has a paid version of their service.
- Shiny application vs. Shiny server
  - Apps a ran locally and use RStudio's service for hosting the app on their servers, on a platform called shinyapps.io
    - \* the free version only allows one to run 5 apps for a certain amount of time per month
    - \* RStudio will send one a message if the limit is reached
    - \* Should one hit the 25-hour per month limit they can send an email to shinyapps-support@rstudio.com to request an increase in their limit so one can continue working on thier project (for this course)
  - A Shiny server is required such that one can host a shiny app for the world
    - \* It requires understanding a little linux server administration and won't be covered in this course.
- Although everything is done in R it is helpful to have some knowledge of HTML to know what the commands in R are actually doing.
- Shiny uses Bootstrap (no relation to the statistical method) style, which suffices for aesthetics, rendering, and resizing to fit different screens.

#### HTML, CSS, and Javascript Tutorials

- Mozilla Developer Network Tutorials
- HTML & CSS from Khan Academy
- Tutorials from Free Code Camp

#### A Shiny Project

- There is a tutorial for shiny on RStudio, however these notes will sort of walk through that tutorial anyway.
- A computer interfaces with the app to generate new plots/results based on the users' input on the app.
- A shiny project consists of a directory with at least two files:
  - ui.R (for user interface) controls how your app looks.
  - server.R that controls what your app does, held on the server.
  - NOTE: The app doesn't actually need these two files but rather the functions the contain, as such it's also an option to just have a app.R file.

- RStudio has some examples of shiny apps too
- I also found someone made a game using Shiny. The game's app also contains a link to the GitHub repo.

#### Shiny Code Demos (1.2)

A demo is located in ./helloShiny

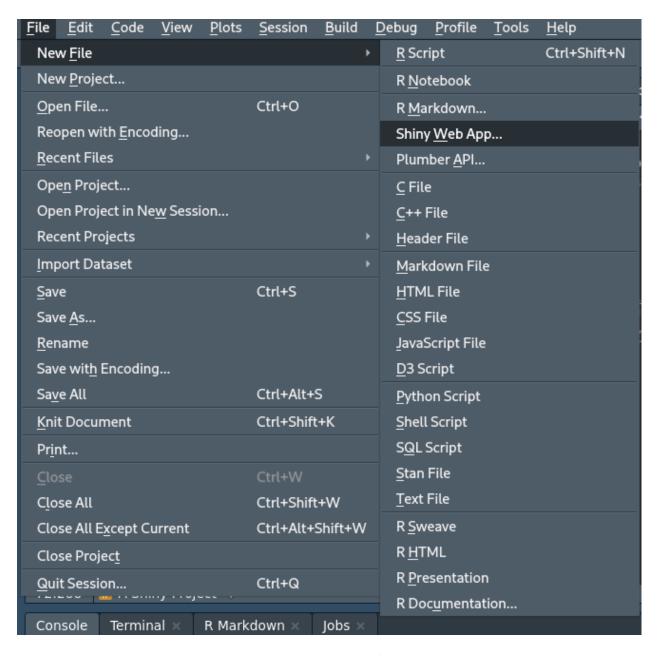


Figure 1: New File

#### Set-Up

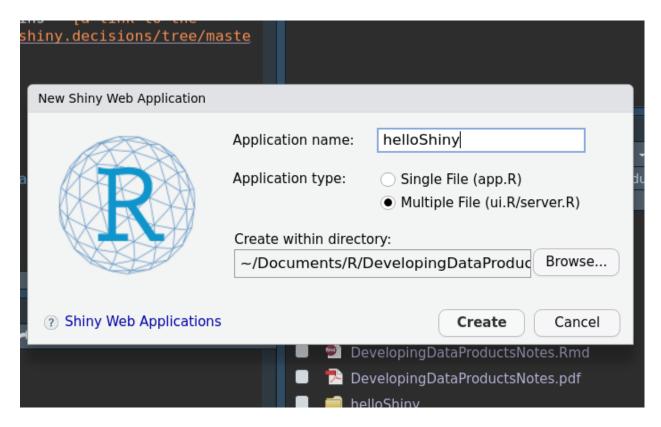


Figure 2: Create app in GUI

#### ui.R

- requires the function shinyUI

#### server.R

- requires the function shinyServer
  - Any logical computations are done here

# Shiny HTML Tags (1.3)

- Shiny provides several wrapper functions for using standard HTML tags in your ui.R, including:
  - h1() through h6() for headlines
  - p() for paragraphs
  - br() for line-breaks
  - a() for inserting hyperlinks
  - div() denotes a section styled with CSS

- span() to color a part of the text
- See the help page, ?builder for more details.
- A guide of HTML tags can be found on this site
- Some of these tags are tested in ./helloShiny/

# Shiny Apps with I/O (1.4)

• In this first demo we'll be looking at a slider input, which will just show the value to the user. This demo is found in ./helloSlider/

# Shiny Apps with Plots (1.5)

- Allowing users to manipulate data nd see the reults of their manipulations as a plot can be very useful
- Shiny provides the plotOutput() function for ui.R
- ...and the renderPlot() function for sever.R
- The demo of this is in ./helloPlot/
- Also contains:
  - in/de-crementing arrows for a textbox
  - checkboxes for a logical if labels are present

#### Shiny Part 2

#### Reactivity (2.1)

- A reactive expression manipulates inputs from Shiny and returns a value.
- Reactivity provides a way for your app to respond, as inputs will change depending on how users interact with your UI
- Expressions that are subject to change should be wrapped with the function reactive()
  - Because of the syntax of Shiny this may look a bit "unorthodox" for R code (Having  $(\{...\})$ )
- Creating a reactive expression is like creating a function:

```
calc_sum <- reactive({
  input$box1 + input$box2
})
# ...
calc_sum()</pre>
```

• Later we'll look at having the user push a button to have the app react, which is helpful for more computational heavy computations.

### reactive Example (2.2)

• An example that uses the reactive function can be found in ./helloReactive/

#### Delayed Reactivity (2.3)

- One may not want an app to immediately react to changes in user input, such as in cases of long-running calculation
- In order to prevent reactive expressions from reacting one can include a submit button in the app.
- An example of this can be found in ./helloSubmit/, which is a "fork" of ./helloReactive/but with the submit button (only comments about the submit button are present in this document).

# Tabs (2.4)

- There are several other kinds of UI components that one can add to an app, such as tabs
  - tabs gives app multiple views
  - navbars to click around multiple tabs, like a top menu
  - sidebars just like navbars but on the side
- There are functions for managing these tabs:
  - tabsetPanel() specifies a group of tabs
  - tabPanel() specifies the contents of an individual tab
- An example of tabs can be found in ./helloTabs/

#### Using your own HTML (2.5)

- To use custom HTML one would save the code as index.html in a directory, named www, which is a child of the directory containing the server.R file
- Most of the style from Shiny's bootstrap setup in the header will no longer be needed

#### Interactive Graphics (2.6)

- A feature of Shiny is the ability to create graphics that a user can interact with.
- One method that can be used to select multiple data points on a graph is by specifying the brush parameter in plotOutput() in the ui.R file, then using the brushedPoints() function in the server.R file.
- The ./helloBrush/ directory gives an example that draws a line of best fit, given user selected points.

# **Shiny Gadgets**

#### Overview (1.1)

- Shiny Gadgets provide a way to use Shiny's interactivity and UI as a part of a data analysis.
- A function is created that opens a small Shiny app in the RStudio viewer pane
  - Since it's smaller we use the miniUI package for manipulating the GUI
  - A gadget is a singular function that contains it's own ui and server functions.
- An example of a gadget to use can be found in myFirstGadget.R (were you expecting helloGadget?), the function's code is displayed below.

```
library(shiny)
library(miniUI)

myFirstGadget <- function() {
    ui <- miniPage(
        gadgetTitleBar("My First Gadget")
    )
    server <- function(input, output, session) {
        # Put in a Done button to close the app
        observeEvent(input$done, {
            stopApp()
        })
    }
    runGadget(ui,server)
}</pre>
```

### Shiny Gadgets (1.2)

- An advantage of Shiny Gadgets is that since they are functions they can take values as arguments and then return values.
- We'll create a simple example of a Gadget that has I/O
  - It will take two different vectors of numbers as arguments & use them to populate two selectInputs. The user can then choose two numbers within the Gadget and their product will be returned
  - (Note: I won't be creating a separate file for this one)

```
library(shiny)
library(miniUI)

multiplyNumbers <- function(numbers1, numbers2) {
  ui <- miniPage(
    gadgetTitleBar("Multiply Two Numbers"),
    miniContentPanel(
    #Make two drop-down boxes of selection options</pre>
```

```
selectInput("num1", "First Number", choices = numbers1),
    selectInput("num2", "Second Number", choices = numbers2)
)

server <- function(input, output, session) {
    observeEvent(input$done, {
            num1 <- as.numeric(input$num1)
            num2 <- as.numeric(input$num2)
            stopApp(num1 * num2)
            })
}

runGadget(ui, server)
}</pre>
```

#### Gadgets with Interactive Graphics (1.3)

- Gadgets can be used to create interactive plots, which can be helpful during the exploratory process or presentation of data.
- Below we'll be looking at an example of this.

```
library(shiny)
library(miniUI)
pickTrees <- function() {</pre>
  #User Interface
 ui <- miniPage(</pre>
    #Title
    gadgetTitleBar("Select Points by Dragging your Mouse"),
    #Show plot with brush interaction
    miniContentPanel(
      plotOutput("plot", #Plot is named "plot"
                 height = "100%", brush = "brush") #Brush is named "brush"
    )
  )
  #Logic
  server <- function(input, output, session) {</pre>
    #Display plot
    output$plot <- renderPlot({</pre>
      plot(trees$Girth, trees$Volume, main = "Trees!",
           xlab = "Girth", ylab = "Volume")
    })
    #On Done button, get selected points & print df of selection to console
    observeEvent(input$done, {
      stopApp(brushedPoints(trees, input$brush,
```

```
xvar = "Girth", yvar = "Volume"))
}

runGadget(ui, server)
}
```

• The original data frame row numbers are retained, as such this can be helpful to quickly identify outliers in one's data

# GoogleVis

# GoogleVis (1.1)

- googleVis is a package that connects R to Google's visualization API
- Allows one to create visuals that would otherwise be difficult to create without it.
- googleVis charts can be embedded into the HTML when using Knitter too
- Btw googleVis uses flash, so unless they change something this will be obsolete after browsers stop supporting flash entirely.
  - Maybe check out the suggested packages in the answer to **this StackOverflow question**.

### Motion Chart Example

#### Charts in googleVis (1.2)

- Motion charts: gvisMotionChart
- Interactive maps: gvisGeoChart
- Interactive tables: gvisTable
- Line charts: gvisLineChart
- Bar charts: gvisColumnChart
- Tree maps: gvisTreeMap

• The full documentation can be fround on CRAN

#### Plots on Maps Example

• Options that can be used with googleVis can be found here (Archived here, just in case)

# Combing Multiple Plots Together

```
M <- gvisMotionChart(Fruits, "Fruit", "Year")

#View locally (setting results = "asis" in Rmd would
# make the HTML code paste right in)
print(M)

# Save HTML to a file
print(M, 'chart', file = "myfilename.html")</pre>
```

#### Viewing the HTML code

#### Things You Can do with Google Vis

- The visualizations can be embedded in websites with HTML code
- Dynamic visualizations can be built with Shiny, Rook, and R.rsp
- Embed them in R markdown based documents
  - Set results="asis" in the chunk options
  - Can be used with kniter and slidify
  - Sometimes clearing kniter cache can help the (static) charts display

#### For More Info

- demo(googleVis) will show off some plots... using flash when needed
- CRAN vignette
- CRAN package info
- Google's documentation
- Google's FAQ

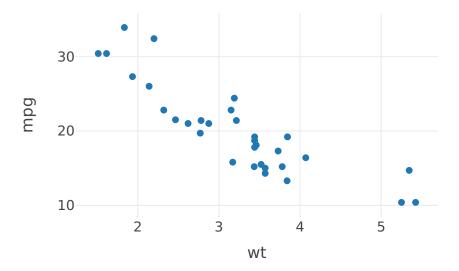
### Plotly

# Plotly Intro (1.1)

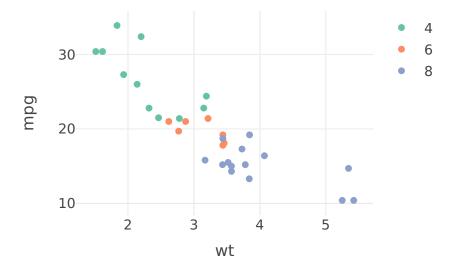
- Plotly is helpful for creating interactive web graphics with R
- Can integrate with a lot of other languages too, such as: R, Python, MatLab, JavaScript, and Microsoft Excel.
- To share visualizations online one will need an account on the plotly site
- However the plotly package will let one just make HTML graphics independent of the plotly site
- Outputs from knitr that warrant for interaction will allow interaction with plotly plots

# Basic Scatter Plot (1.2)

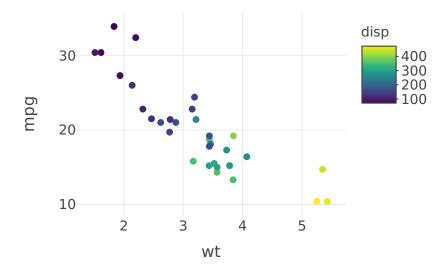
• A scatter plot is indicated with the parameter mode = "markers"



# Adding Qualatative Variables (1.3)



# Discreet Color



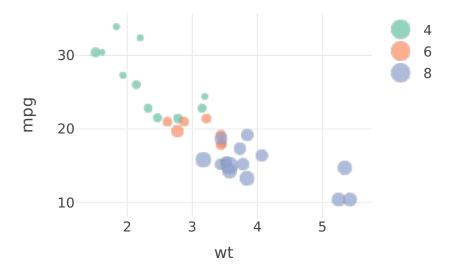
# **Continuous Color**

# Altering Size

## Warning: 'line.width' does not currently support multiple values.

## Warning: 'line.width' does not currently support multiple values.

## Warning: 'line.width' does not currently support multiple values.



# 3D Scatter Plot

- A three-dimensional scatterplot can be created with the type = "scatter3d" argument.
   These plots can be clicked and dragged to view different angles
- Uses WebGL to create interactive 3D graphics in a web page

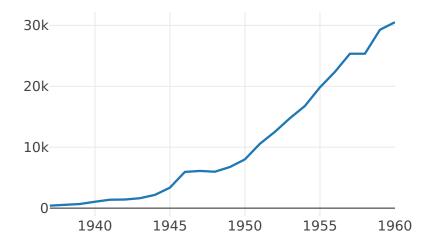
WebGL is not supported by your browser visit https://get.webgl.org for more info

# Line Graphs (1.4)

# Standard Line Graph Example

• Default for plot\_ly is to make a scatter plot, so the argument mode = "line" has to be passed.

```
data("airmiles")
head(cbind(time(airmiles), airmiles))
##
        time(airmiles) airmiles
## [1,]
                  1937
                             412
## [2,]
                             480
                   1938
## [3,]
                  1939
                             683
## [4,]
                  1940
                            1052
## [5,]
                            1385
                  1941
## [6,]
                  1942
                            1418
plot_ly(x = time(airmiles), y = airmiles, mode = "line")
```



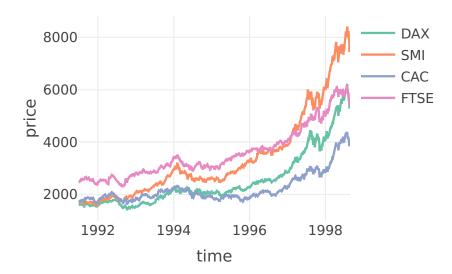
# Multi Line Graph - Stock Example

• Multiple lines can be shown by specifying the column in the data frame that separates the lines

```
## variable price index time
## 1 DAX 1628.75 DAX 1991.496
## 2 DAX 1613.63 DAX 1991.500
## 3 DAX 1606.51 DAX 1991.504
## 4 DAX 1621.04 DAX 1991.508
```

```
## 5     DAX 1618.16     DAX 1991.512
## 6     DAX 1610.61     DAX 1991.515

plot_ly(stocks, x = ~time, y = ~price, color = ~index, mode = "line")
```

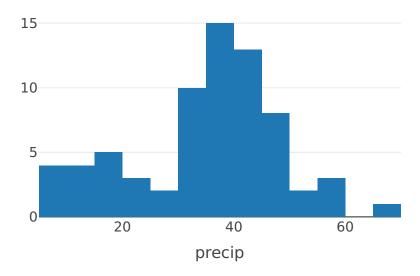


# Other Plots (1.5)

# Histogram

• Use the type = "histogram" argument

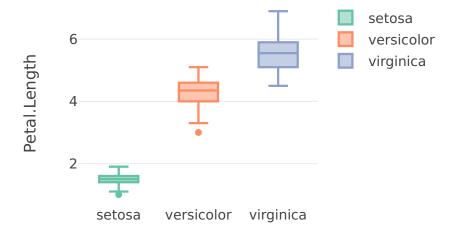
plot\_ly(x = ~precip, type = "histogram")



# Boxplot

• Use the type = "box" argument

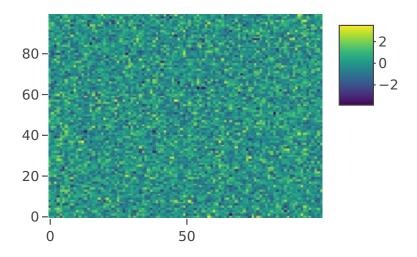
```
plot_ly(iris, y = "Petal.Length, color = "Species, type = "box")
```



# Heatmap

• Use the type = "heatmap" argument

```
set.seed(1618033)
terrain1 <- matrix(rnorm(100*100), nrow = 100, ncol = 100)
plot_ly(z = terrain1, type = "heatmap")</pre>
```



# 3D Surface

• Create move-able 3D surfaces with type = "surface"

```
set.seed(1618)
terrain2 <- matrix(sort(rnorm(100*100)), nrow = 100, ncol = 100)
plot_ly(z = terrain2, type = "surface")</pre>
```

WebGL is not supported by your browser visit https://get.webgl.org for more info

# Choropleth Maps (1.6)

• Geographical maps with regions colored by a variable

```
# Create data frame
state_pop <- data.frame(State = state.abb, Pop = as.vector(state.x77[,1]))</pre>
#Create hover text
state_pop$hover <- with(state_pop, paste(State, "<br>", "Population:", Pop))
head(state_pop)
    State Pop
##
                                     hover
## 1 AL 3615 AL <br > Population: 3615
## 2 AK 365 AK <br > Population: 365
## 3 AZ 2212 AZ <br > Population: 2212
## 4 AR 2110 AR <br > Population: 2110
## 5
       CA 21198 CA <br > Population: 21198
       CO 2541 CO <br > Population: 2541
#Make state borders red
borders <- list(color = toRGB("#FF0000"))</pre>
#Set up some mapping options
map_options <- list(</pre>
```

# US Population in 1975



# Displaying ggplots with Plotly (1.7)

```
set.seed(100)
d <- diamonds[sample(nrow(diamonds), 1000), ]
p <- ggplot(data = d, aes(x = carat, y = price)) +
    geom_point(aes(text = paste("Clarity:", clarity)), size = 4) +
    geom_smooth(aes(colour = cut, fill = cut)) + facet_wrap(~ cut)</pre>
```

```
## Warning: Ignoring unknown aesthetics: text

(gg <- ggplotly(p))

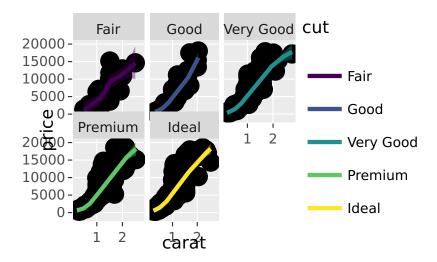
## Warning: 'group_by_()' is deprecated as of dplyr 0.7.0.

## Please use 'group_by()' instead.

## See vignette('programming') for more help

## This warning is displayed once every 8 hours.

## Call 'lifecycle::last_warnings()' to see where this warning was generated.</pre>
```



#### Posting to Plotly (1.8)

• With a saved plotly plot, gg in this case, one wants to execute plotly\_POST(gg), however you have to have an authentication API established:

```
Sys.setenv("plotly_username" = "your_plotly_username")
Sys.setenv("plotly_api_key" = "your_api_key")
```

- One can save these commands in their .Rprofile file to run everytime R is started
- Extra note: For kniting PDFs one can see this StackOverflow post for instructions

#### Reminder to Commit (05), Delete this line AFTER Committing

#### Quiz 1

- 1. Which of the following are absolutely necessary for creating a functioning shiny app?
- A ui.R file containing a call to shinyUI()
- A server.R file containing a call to shinyServer()
- 2. What is incorrect about the following syntax in ui.R?

```
library(shiny)
shinyUI(pageWithSidebar(
  headerPanel("Data science FTW!"),
  sidebarPanel(
    h2('Big text')
    h3('Sidebar')
),
  mainPanel(
    h3('Main Panel text')
)
```

- Missing a comma in the sidebar panel
- 3. Consider the following in ui.R

And the following in server.R

```
text(63, 140, paste("MSE = ", round(mse, 2)))
})
}
```

- The server.R output name isn't the same as the plotOutput command used in ui.R
- 4. What are the main differences between creating a Shiny Gadget and creating a regular Shiny App?
- Shiny Gadgets are designed to be used by R users in the middle of a data analysis
- Shiny Gadgets are designed to have small user interfaces that fit on one page
- 5. Consider the following R script

```
library(shiny)
library(miniUI)
pickXY <- function() {</pre>
 ui <- miniPage(</pre>
    gadgetTitleBar("Select Points by Dragging your Mouse"),
    miniContentPanel(
      plotOutput("plot", height = "100%", brush = "brush")
    )
  )
  server <- function(input, output, session) {</pre>
      output$plot <- renderPlot({</pre>
        plot(data_frame$X, data_frame$Y, main = "Plot of Y versus X",
           xlab = "X", ylab = "Y")
      })
      observeEvent(input$done, {
        stopApp(brushedPoints(data_frame, input$brush,
                            xvar = "X", yvar = "Y"))
      })
  }
 runGadget(ui, server)
my_data <- data.frame(X = rnorm(100), Y = rnorm(100))</pre>
pickXY(my_data)
```

• No arguments are defined for pickXY()

# R Markdown and Leaflet

#### R. Markdown

• Personal note: I've taken notes on R Markdown in a previous course and have linked back to it whenever the topic came up, however having it in a PDF format may be nice for later so I'll be repeating some material previously covered in past courses.

# Overview (1.1)

- R Markdown is built into RStudio and allows one to create documents (like HTML, PDF, and Word) from R.
- R Markdown allows one to embed R code & plots into documents.
- It allows work in R to be in more accessible formats for the laymen
- R Markdown documents are reproducible the source code gets ran every time a doc. is generated, so if data change or source code changes, the output will reflect these changes.
  - Reproducablity helps increase credibility for a given report (making sure results aren't fake).
- Helpful for generating routine reports, especially ones based on a particular set of data

### Set Up - Presentaion (1.2)

- Create a new R Markdown file in RStudio by selecting File > New File > R Markdown...
- Select "presentation" tab
- Enter a title, author, and select what kind of output you want (these options can be changed later)
- An example of the default template's, ioslides\_presentation, R Markdown and HTML output can be found in ./helloPresentation/.

#### Editing a Presentaion (1.3 - 1.5)

Example & output is in ./editingPresentaions

#### Publishing to Github Pages (1.6)

- A guide to making github pages can be found here.

  If one has the github link of: https://github.com/seankross/slides/blob/gh-pages/Developing\_
  Data\_Products/R\_Markdown/R\_Markdown.html
- Where:
  - seankross is the user
  - slides is the repo

- The github pages can be found by going to .github.io//
  - In this example that would be  $https://seankross.github.io/slides/Developing\_Data\_Products/R\_Markdown/R\_Markdown.html$
  - Note that "blob/gh-pages" is removed as this is internal info for GitHub's main site that a gh-pages exists

# Sharing R Markdown Documents

- R Markdown documents can be shared with:
  - GitHub Pages
  - RPubs
    - \* RPubs Instructions

# NeoCities - for hosting HTML files for free Reminder to Commit (06), Delete this line AFTER Committing Leaflet Leaflet (1.1) Leaflet (1.2) Leaflet (1.3) Leaflet (1.4) Leaflet (1.5) Leaflet (1.6) Reminder to Commit (07), Delete this line AFTER Committing Quiz 2 1. 2. 3. 4. 5. 6.

Reminder to Commit (Q2), Delete this line AFTER Committing

# Course Project 1

Reminder to Commit (P1), Delete this line AFTER Committing

# Building R Packages

R Packages
R Packages Overview
R Packages (1.1)
R Packages (1.2)
Building R Packages Demo
Reminder to Commit (08), Delete this line AFTER Committing
R Classes and Methods
R Classes and Methods (1.1)
R Classes and Methods (1.2)
Reminder to Commit (09), Delete this line AFTER Committing
Quiz 3
1.
2.
3.
4.
Reminder to Commit (Q3), Delete this line AFTER Committing
Course Project 2
Reminder to Commit (P2), Delete this line AFTER Committing
Swirl
Swirl
Swirl (1.1)
Swirl (1.2)
Swirl (1.3)
Reminder to Commit (10), Delete this line AFTER Committing
Course Project 3
Reminder to Commit (P3), Delete this line BEFORE Committing