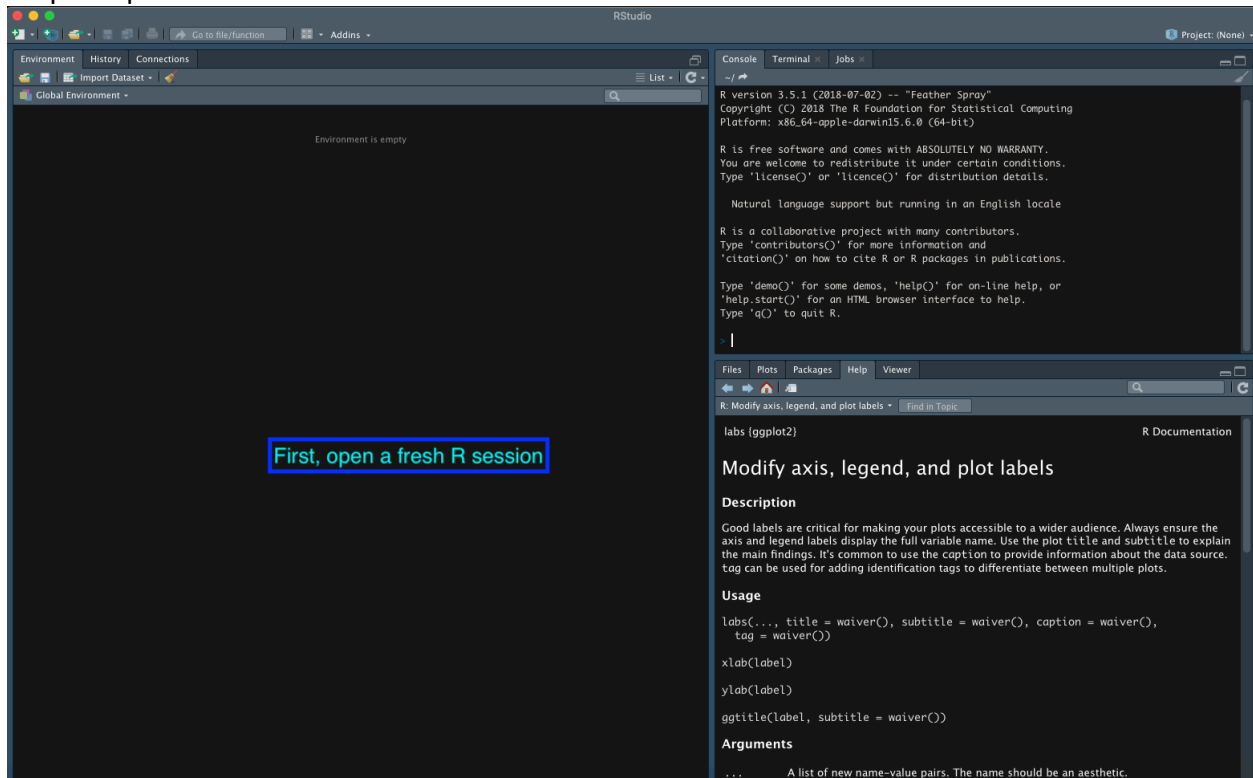
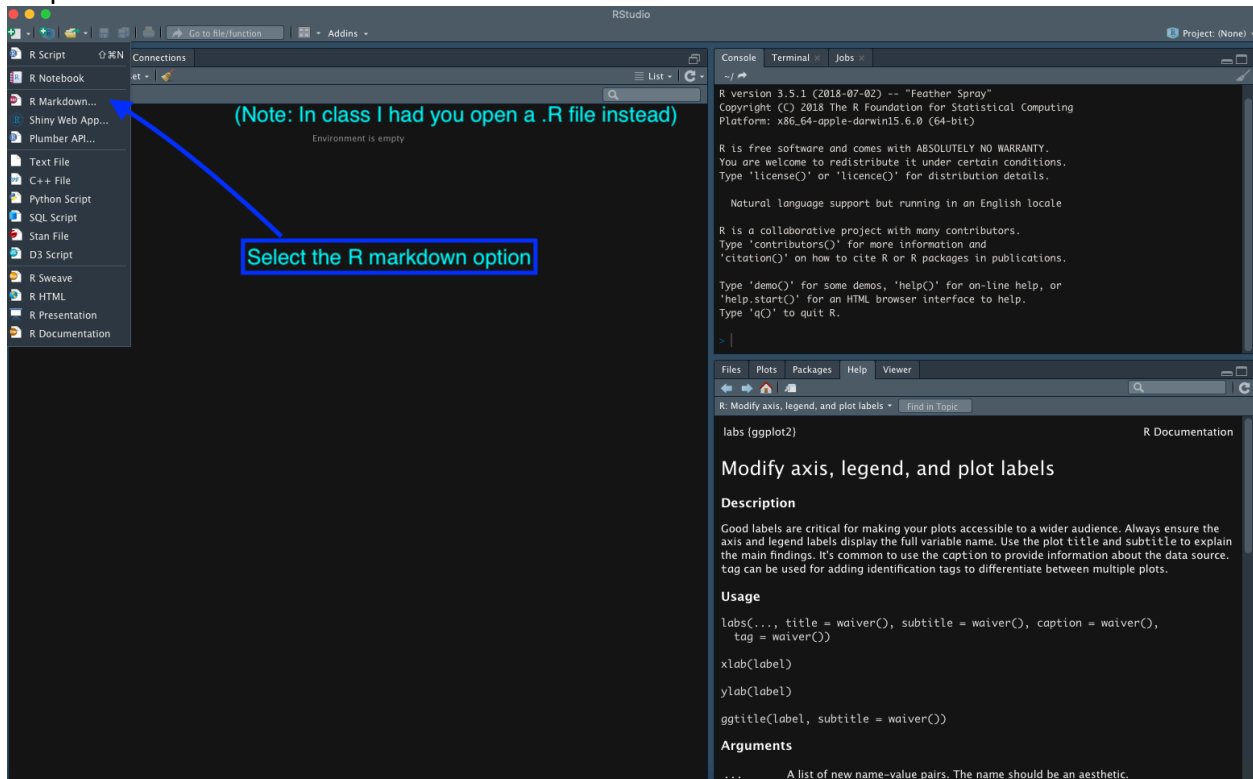


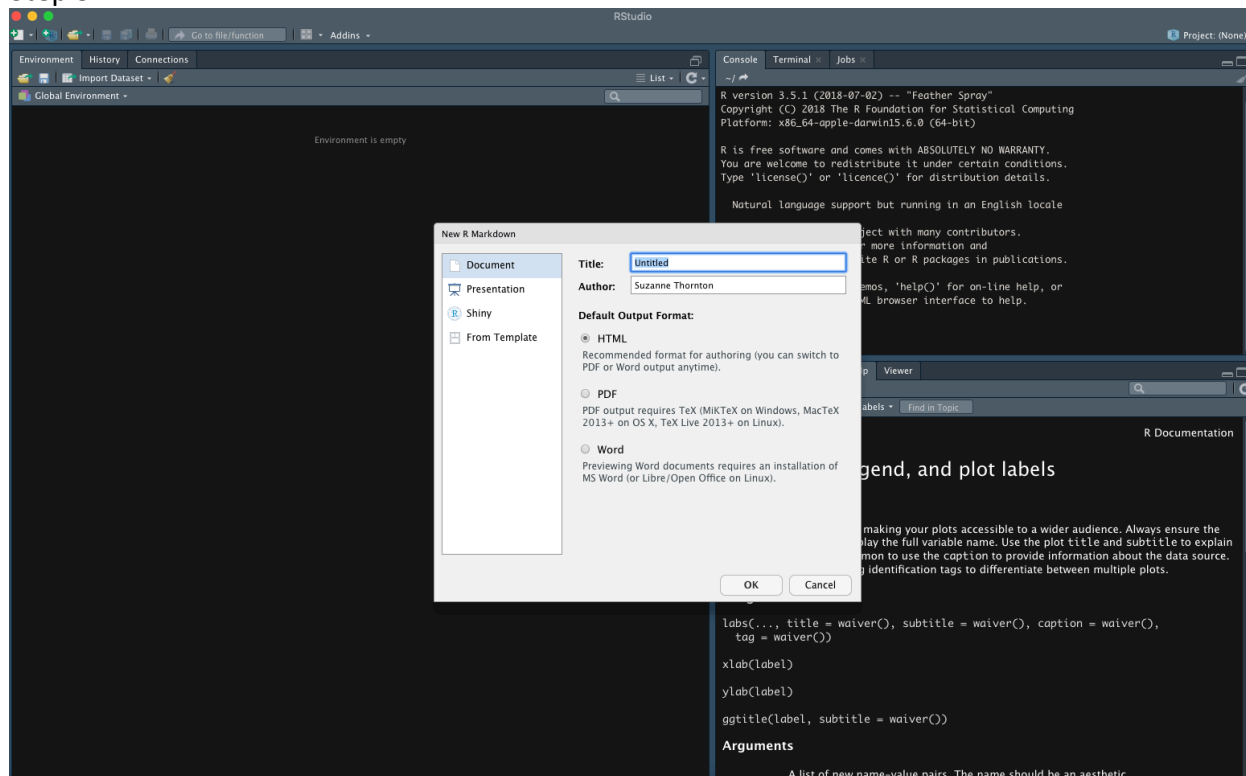
## Step 1: Open R Studio



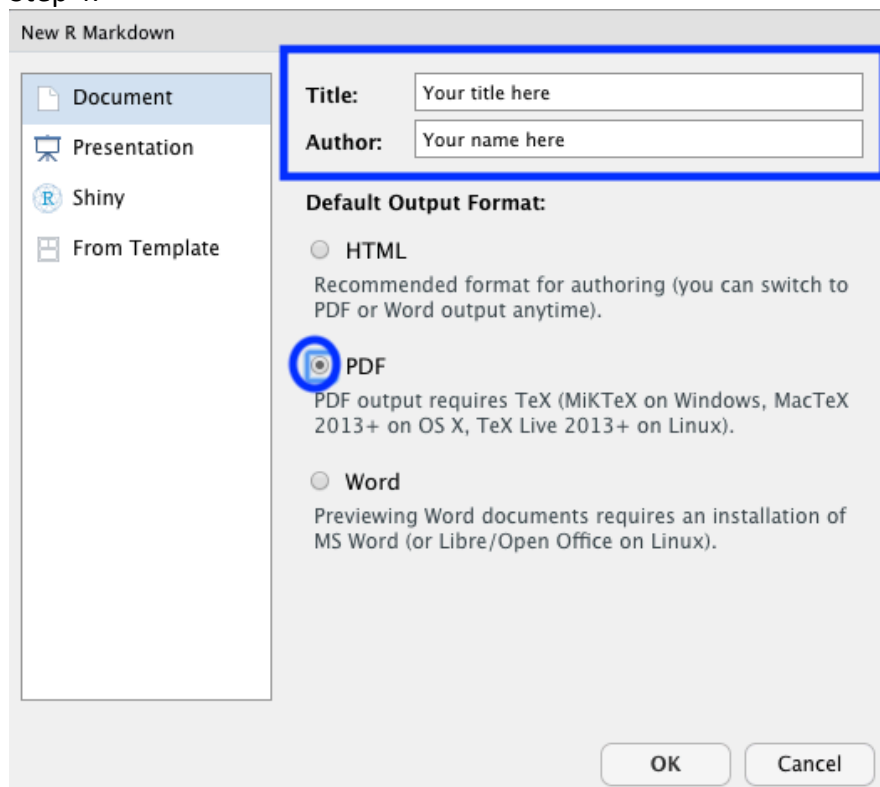
## Step 2:



### Step 3:



### Step 4:



## Step 5:

R will automatically create a .Rmd template file. We won't use this however.

```
1 ---
2 title: "Your title here"
3 author: "Your name here"
4 date: "9/6/2019"
5 output: pdf_document
6 ---
7
8 {r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word
15 documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 When you click the "Knit" button a document will be generated that includes both content as well as the output
18 of any embedded R code chunks within the document. You can embed an R code chunk like this:
19
20 {r cars}
21 summary(cars)
22
23 ## Including Plots
24
25 You can also embed plots, for example:
26
27 {r pressure, echo=FALSE}
28 plot(pressure)
29
30 Note that the 'echo = FALSE' parameter was added to the code chunk to prevent printing of the R code that
31 generated the plot.
```

R version 3.5.1 (2018-07-02) -- "Feather Spray"  
Copyright (C) 2018 The R Foundation for Statistical Computing  
Platform: x86\_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

Files Plots Packages Help Viewer

R. Modify axis, legend, and plot labels Find in Topic

labs(ggplot2)

### Modify axis, legend, and plot labels

#### Description

Good labels are critical for making your plots accessible to a wider audience. Always ensure the axis and legend labels display the full variable name. Use the plot title and subtitle to explain the main findings. It's common to use the caption to provide information about the data source. tag can be used for adding identification tags to differentiate between multiple plots.

#### Usage

```
labs(..., title = waiver(), subtitle = waiver(), caption = waiver(),
tag = waiver())
```

```
xlab(label)
ylab(label)
ggtitle(label, subtitle = waiver())
```

#### Arguments

... A list of new name-value pairs. The name should be an aesthetic.

## Step 6:

So highlight and erase all of the default text

```
1 ---
2 title: "Your title here"
3 author: "Your name here"
4 date: "9/6/2019"
5 output: pdf_document
6 ---
7
8 {r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word
15 documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 When you click the "Knit" button a document will be generated that includes both content as well as the output
18 of any embedded R code chunks within the document. You can embed an R code chunk like this:
19
20 {r cars}
21 summary(cars)
22
23 ## Including Plots
24
25 You can also embed plots, for example:
26
27 {r pressure, echo=FALSE}
28 plot(pressure)
29
30 Note that the 'echo = FALSE' parameter was added to the code chunk to prevent printing of the R code that
31 generated the plot.
```

R version 3.5.1 (2018-07-02) -- "Feather Spray"  
Copyright (C) 2018 The R Foundation for Statistical Computing  
Platform: x86\_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.  
You are welcome to redistribute it under certain conditions.  
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.  
Type 'contributors()' for more information and  
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or  
'help.start()' for an HTML browser interface to help.  
Type 'q()' to quit R.

Files Plots Packages Help Viewer

R. Modify axis, legend, and plot labels Find in Topic

labs(ggplot2)

### Modify axis, legend, and plot labels

#### Description

Good labels are critical for making your plots accessible to a wider audience. Always ensure the axis and legend labels display the full variable name. Use the plot title and subtitle to explain the main findings. It's common to use the caption to provide information about the data source. tag can be used for adding identification tags to differentiate between multiple plots.

#### Usage

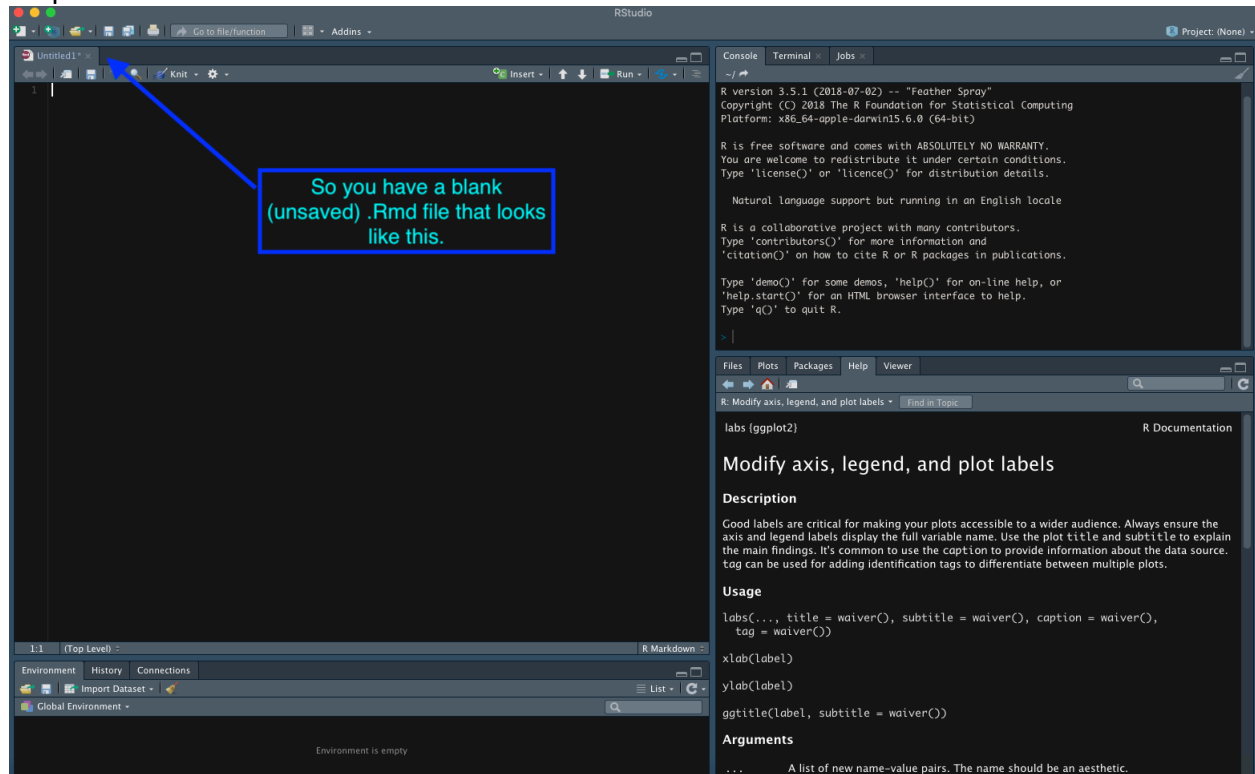
```
labs(..., title = waiver(), subtitle = waiver(), caption = waiver(),
tag = waiver())
```

```
xlab(label)
ylab(label)
ggtitle(label, subtitle = waiver())
```

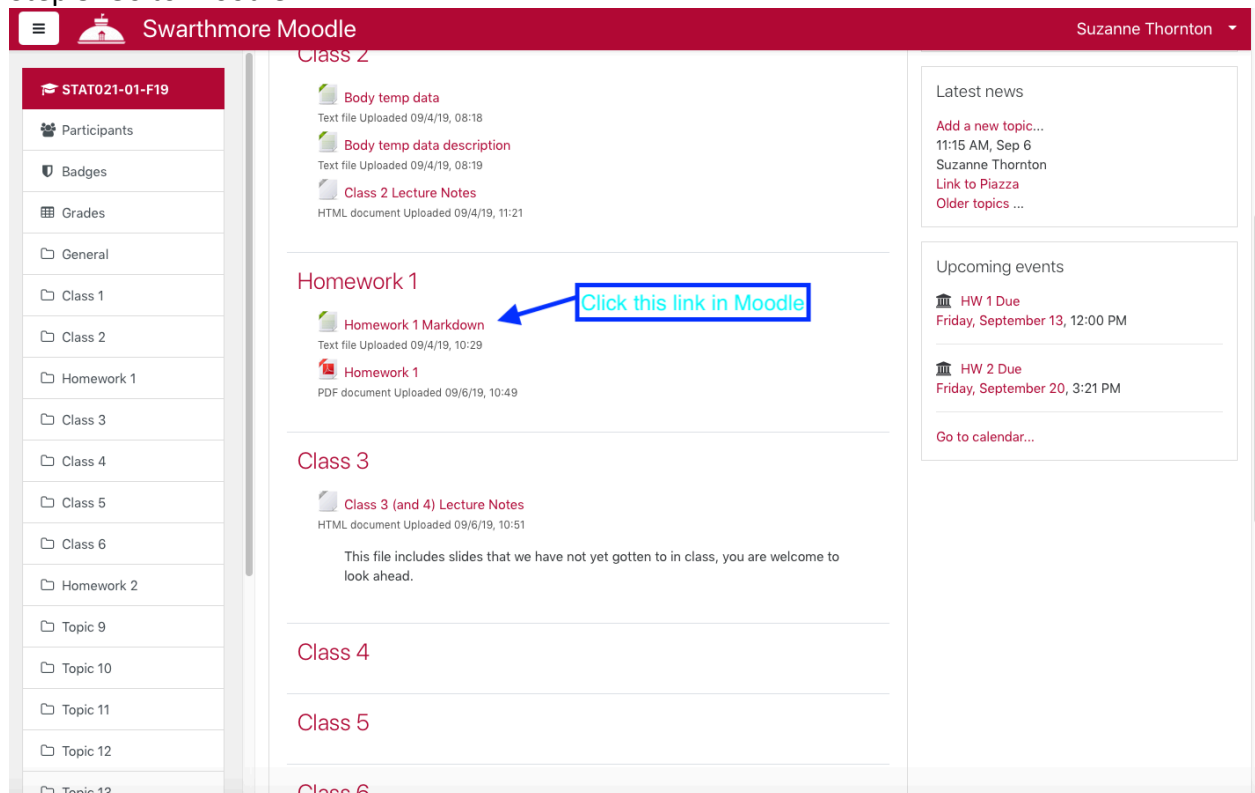
#### Arguments

... A list of new name-value pairs. The name should be an aesthetic.

## Step 7:



## Step 8: Go to Moodle



## Step 9:

```
---
title: "Stat 021 Homework 1"
author: "Suzanne Thornton"
date: "Due: Fri, Sept 13"
output: pdf_document
---
```

This will open up a new tab that looks like this

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse) ## includes ggplot2, dplyr, tidyr, readr, purrr, tibble, stringr, forcats
library(magrittr) ## for piping
#library(glue) ## for easier string and variable concatenation
```
```

<center>\*\*\*Instructions:\*\*\* A hard copy of your homework must be handed in to me at the end of class on the due date or I must have recieved via email a \*\*pdf\*\* version of your homework by \*\*noon\*\* on the due date. If you are having trouble getting your \*.Rmd\* file to compile, you need to get help with this \*\*before\*\* the due date.

You are allowed to hand in \*\*only one\*\* late homework assignment throughout the semester. If you need to hand in this particular assignment late, you must let me know via email by noon on the due date.

You are encouraged to study with your peers to help complete the homework assignments but no copying is allowed. If I see that two or more homework assignments are copied, all students involved will recieve a grade of \$0\$ on that assignment and will forfeit (perhaps retroactively) the opportunity to hand in a late homework. </center>

\*\*\*

\*\*Q 1)\*\* Download and install R and R Studio following the instructions in class. Install the package \*swirl()\* using the command "install.packages("swirl")". Once the package is installed, call the package to your R session using the command "library("swirl")". Follow the instructions that pop up in your console. Select the course option "1: R Programming: The basics of programming in R" and complete the following lessons (about \$20\$ minutes to complete each)

- 1: Basic Building Blocks
- 2: Workspace and Files
- 8: Logic

Even though you don't get points for doing this, it will really help you to learn how to program in R. If the tutorial is covering something that you already know how to do, use the \*skip()\* command to move through the tutorial faster, but note that there are some questions which you will not be able to skip and you'll be forced to think through. (\$0\$ points)

\*\*Q 2)\*\* Design your own experimental study (made up) on a population of your choice. Provide details on what is the population under study, how will you collect a sample, what are potential biases keeping your sample from truely being random and define different treatment/control groups and describe how you will randomly assign treatments to your observational units. (\$5\$ points)

\*\*Q 3)\*\*

Access the data set called \*sleep\* in R. Note this data set is loaded into R automatically (you do not need to import it or install a package to access it). (\$10\$ points)

1. Define a new variable called \*group1.sleep\* that includes only the values of the variable \*extra\* for individuals from group 1.

1. Perform a t-test on ``group1.sleep`` to test if the extra hours slept by group 1 is smaller than or equal to \$0.5\$ hours at an  $\alpha = 0.1$  level. I.e. Test the hypothesis  $H_0: \mu = 0.5$  vs.  $H_1: \mu \neq 0.5$  at an  $\alpha = 0.1$  significance level.

1. Report and interpret the \$90\%\$ CI for the average extra hours of sleep for group 1.

## Step 10:

```
---
title: "Stat 021 Homework 1"
author: "Suzanne Thornton"
date: "Due: Fri, Sept 13"
output: pdf_document
---
```

Highlight all of the text on this page and copy/paste it into your .Rmd file

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(tidyverse) ## includes ggplot2, dplyr, tidyr, readr, purrr, tibble, stringr, forcats
library(magrittr) ## for piping
#library(glue) ## for easier string and variable concatenation
```
```

<center>\*\*\*Instructions:\*\*\* A hard copy of your homework must be handed in to me at the end of class on the due date or I must have recieved via email a \*\*pdf\*\* version of your homework by \*\*noon\*\* on the due date. If you are having trouble getting your \*.Rmd\* file to compile, you need to get help with this \*\*before\*\* the due date.

You are allowed to hand in \*\*only one\*\* late homework assignment throughout the semester. If you need to hand in this particular assignment late, you must let me know via email by noon on the due date.

You are encouraged to study with your peers to help complete the homework assignments but no copying is allowed. If I see that two or more homework assignments are copied, all students involved will recieve a grade of \$0\$ on that assignment and will forfeit (perhaps retroactively) the opportunity to hand in a late homework. </center>

\*\*\*

\*\*Q 1)\*\* Download and install R and R Studio following the instructions in class. Install the package \*swirl()\* using the command "install.packages("swirl")". Once the package is installed, call the package to your R session using the command "library("swirl")". Follow the instructions that pop up in your console. Select the course option "1: R Programming: The basics of programming in R" and complete the following lessons (about \$20\$ minutes to complete each)

- 1: Basic Building Blocks
- 2: Workspace and Files
- 8: Logic

Even though you don't get points for doing this, it will really help you to learn how to program in R. If the tutorial is covering something that you already know how to do, use the \*skip()\* command to move through the tutorial faster, but note that there are some questions which you will not be able to skip and you'll be forced to think through. (\$0\$ points)

\*\*Q 2)\*\* Design your own experimental study (made up) on a population of your choice. Provide details on what is the population under study, how will you collect a sample, what are potential biases keeping your sample from truely being random and define different treatment/control groups and describe how you will randomly assign treatments to your observational units. (\$5\$ points)

\*\*Q 3)\*\*

Access the data set called \*sleep\* in R. Note this data set is loaded into R automatically (you do not need to import it or install a package to access it). (\$10\$ points)

1. Define a new variable called \*group1.sleep\* that includes only the values of the variable \*extra\* for individuals from group 1.

1. Perform a t-test on ``group1.sleep`` to test if the extra hours slept by group 1 is smaller than or equal to \$0.5\$ hours at an  $\alpha = 0.1$  level. I.e. Test the hypothesis  $H_0: \mu = 0.5$  vs.  $H_1: \mu \neq 0.5$  at an  $\alpha = 0.1$  significance level.

1. Report and interpret the \$90\%\$ CI for the average extra hours of sleep for group 1.

The screenshot displays the RStudio interface. The main editor window shows an R Markdown file with a code chunk containing R code for setting up the environment and installing packages. A blue rectangular box is drawn over the text "So now your .Rmd file should look exactly like this", which is overlaid on the R code. The console on the right shows the R version (3.5.1) and the R Foundation for Statistical Computing logo. The environment pane at the bottom shows the "Stat 021 Homework 1" project.

[illegible]



## Step 13:

Save As: Lastname\_firstname\_HW1

Tags:

Stat 21

Please save your document using this format for the title (you will be submitting your output pdf files online)

RStudio

R version 3.5.1 (2018-07-02) -- "Feather Spray"

Console

Terminal

Jobs

Project: (None)

Environment

History

Connections

Global Environment

Environment is empty

Stat 021 Homework 1

Instructions: A hard copy of your homework must be handed in to me at the end of class on the due date. If I must have received your final file to compile, you need to get help with this before the due date.

You are allowed to hand in only one late homework assignment throughout the semester. If you need to hand in the particular assignment late, you must let me know via email by noon on the due date.

You are encouraged to study with your peers to help complete the homework assignments but no copying is allowed. If I see that two or more homework assignments are copied, all students involved will receive a grade of 0 on that assignment and will forfeit (perhaps retroactively) the opportunity to hand in a late homework.

• 1: Basic Building Blocks

• 2: Workspace and Files

Q 1) Download and install R and R Studio following the instructions in class. Install the package 'swirl' using the command "install.packages('swirl')". Once the package is installed, call the package to your R session using the command "library('swirl')". Follow the instructions that pop up in your console. Select the course option "1: R Programming: The basics of programming in R" and complete the following lessons (about 20 minutes to complete each):

• 1: Basic Building Blocks

• 2: Workspace and Files

• 3: Logic

Even though you don't get points for doing this, it will really help you to learn how to program in R. If the material is covering something that you already know how to do, use the skip() command to move through the tutorial faster, but note that there are some questions which you will not be able to skip and you'll be forced to think through. (4 points)

Q 2) Design your own experimental study (made up) on a population of your choice. Provide details on what is the population under study, how will you collect a sample, what are potential biases keeping your sample from truly being random and define different treatment/control groups and describe how you will randomly assign treatments to your observational units. (5 points)

Q 3) Across the data set called sleep in R. Note this data set is loaded into R automatically (you do not need to import it or install a package to access it). (20 points)

1. Define a new variable called group.sleep that includes only the values of the variable extra.hours for individuals from group 1.

2. Perform a t-test on "group.sleep" to test if the extra hours slept by group 1 is smaller than or equal to 0.5 hours at an  $\alpha = 0.1$  level. I.e. Test the hypothesis  $H_0: \mu \geq 0.5$  vs.  $H_1: \mu < 0.5$  at an  $\alpha = 0.1$  significance level.

3. Report and interpret the 90% CI for the average extra hours of sleep for group 1.

4. Form a new categorical variable called extra.hours that categorizes the variable extra.hours into two groups, the first where extra.hours sleep is at least one hour and the second where extra.hours sleep is less than one hour. Print a table that counts the total number of observations in each group. (You may want to use the table() function.)

5. Produce two boxplots for the variable extra.hours, one corresponding to each group. Make sure each plot has

plot labels

Description

Good labels are critical for making your plots accessible to a wider audience. Always ensure the axis and legend labels display the full variable name. Use the plot title and subtitle to explain the main findings. It's common to use the caption to provide information about the data source. tag can be used for adding identification tags to differentiate between multiple plots.

Usage

labs(..., title = waiver(), subtitle = waiver(), caption = waiver(), tag = waiver())

xlab(label)

ylib(label)

ggtitle(label, subtitle = waiver())

Arguments

... A list of new name-value pairs. The name should be an aesthetic.

## Step 14:

Stat 021 Homework 1

Suzanne Thornton

Due: Fri, Sept 13

Instructions: A hard copy of your homework must be handed in to me at the end of class on the due date. If I must have received your final file to compile, you need to get help with this before the due date.

You are allowed to hand in only one late homework assignment throughout the semester. If you need to hand in the particular assignment late, you must let me know via email by noon on the due date.

You are encouraged to study with your peers to help complete the homework assignments but no copying is allowed. If I see that two or more homework assignments are copied, all students involved will receive a grade of 0 on that assignment and will forfeit (perhaps retroactively) the opportunity to hand in a late homework.

• 1: Basic Building Blocks

• 2: Workspace and Files

• 3: Logic

Even though you don't get points for doing this, it will really help you to learn how to program in R. If the material is covering something that you already know how to do, use the skip() command to move through the tutorial faster, but note that there are some questions which you will not be able to skip and you'll be forced to think through. (4 points)

Q 2) Design your own experimental study (made up) on a population of your choice. Provide details on what is the population under study, how will you collect a sample, what are potential biases keeping your sample from truly being random and define different treatment/control groups and describe how you will randomly assign treatments to your observational units. (5 points)

Q 3) Across the data set called sleep in R. Note this data set is loaded into R automatically (you do not need to import it or install a package to access it). (20 points)

1. Define a new variable called group.sleep that includes only the values of the variable extra.hours for individuals from group 1.

2. Perform a t-test on "group.sleep" to test if the extra hours slept by group 1 is smaller than or equal to 0.5 hours at an  $\alpha = 0.1$  level. I.e. Test the hypothesis  $H_0: \mu \geq 0.5$  vs.  $H_1: \mu < 0.5$  at an  $\alpha = 0.1$  significance level.

3. Report and interpret the 90% CI for the average extra hours of sleep for group 1.

4. Form a new categorical variable called extra.hours that categorizes the variable extra.hours into two groups, the first where extra.hours sleep is at least one hour and the second where extra.hours sleep is less than one hour. Print a table that counts the total number of observations in each group. (You may want to use the table() function.)

5. Produce two boxplots for the variable extra.hours, one corresponding to each group. Make sure each plot has

plot labels

terminal

R Markdown

Jobs

Note the three different tabs in this window. One is the R terminal, one for R Markdown messages and the Jobs tab we won't worry about.

This is the output pdf document once the file has been knitted

Environment

History

Connections

Global Environment

Environment is empty

Stat 021 Homework 1

Instructions: A hard copy of your homework must be handed in to me at the end of class on the due date. If I must have received your final file to compile, you need to get help with this before the due date.

You are allowed to hand in only one late homework assignment throughout the semester. If you need to hand in the particular assignment late, you must let me know via email by noon on the due date.

You are encouraged to study with your peers to help complete the homework assignments but no copying is allowed. If I see that two or more homework assignments are copied, all students involved will receive a grade of 0 on that assignment and will forfeit (perhaps retroactively) the opportunity to hand in a late homework.

• 1: Basic Building Blocks

• 2: Workspace and Files

• 3: Logic

Even though you don't get points for doing this, it will really help you to learn how to program in R. If the material is covering something that you already know how to do, use the skip() command to move through the tutorial faster, but note that there are some questions which you will not be able to skip and you'll be forced to think through. (4 points)

Q 2) Design your own experimental study (made up) on a population of your choice. Provide details on what is the population under study, how will you collect a sample, what are potential biases keeping your sample from truly being random and define different treatment/control groups and describe how you will randomly assign treatments to your observational units. (5 points)

Q 3) Across the data set called sleep in R. Note this data set is loaded into R automatically (you do not need to import it or install a package to access it). (20 points)

1. Define a new variable called group.sleep that includes only the values of the variable extra.hours for individuals from group 1.

2. Perform a t-test on "group.sleep" to test if the extra hours slept by group 1 is smaller than or equal to 0.5 hours at an  $\alpha = 0.1$  level. I.e. Test the hypothesis  $H_0: \mu \geq 0.5$  vs.  $H_1: \mu < 0.5$  at an  $\alpha = 0.1$  significance level.

3. Report and interpret the 90% CI for the average extra hours of sleep for group 1.

4. Form a new categorical variable called extra.hours that categorizes the variable extra.hours into two groups, the first where extra.hours sleep is at least one hour and the second where extra.hours sleep is less than one hour. Print a table that counts the total number of observations in each group. (You may want to use the table() function.)

5. Produce two boxplots for the variable extra.hours, one corresponding to each group. Make sure each plot has

plot labels

Description

Good labels are critical for making your plots accessible to a wider audience. Always ensure the axis and legend labels display the full variable name. Use the plot title and subtitle to explain the main findings. It's common to use the caption to provide information about the data source. tag can be used for adding identification tags to differentiate between multiple plots.

Usage

labs(..., title = waiver(), subtitle = waiver(), caption = waiver(), tag = waiver())

xlab(label)

ylib(label)

ggtitle(label, subtitle = waiver())

Arguments

... A list of new name-value pairs. The name should be an aesthetic.