Practice

Tyson S. Barrett

6/22/2017

# This is Our Practice Document

## Section 1

What do you think all these do?

## 1.   
library(tidyverse)  
  
## 2.  
setwd("~/Dropbox/Teaching/R for Social Sciences/Data/")  
  
## 3.  
d = read.csv("WideFormat\_TheOffice.csv") %>%  
 mutate(Sex = factor(Sex))  
  
## 4.  
summary(d)

:) X Name Prod1 MentalApt PhysApt   
:) Min. : 1.0 Andy :1 Min. :1.0 Min. :1.0 Min. :2.000   
:) 1st Qu.: 4.5 Angela :1 1st Qu.:2.5 1st Qu.:4.0 1st Qu.:4.000   
:) Median : 8.0 Creed :1 Median :3.0 Median :5.0 Median :6.000   
:) Mean : 8.0 Dwight :1 Mean :3.2 Mean :5.2 Mean :5.733   
:) 3rd Qu.:11.5 Jim :1 3rd Qu.:4.0 3rd Qu.:7.0 3rd Qu.:7.000   
:) Max. :15.0 Kelley :1 Max. :5.0 Max. :8.0 Max. :8.000   
:) (Other):9   
:) Married Sex Race Income   
:) Min. :0.0000 0:10 Black : 1 Min. :35.00   
:) 1st Qu.:0.0000 1: 5 Indian : 1 1st Qu.:42.50   
:) Median :0.0000 Mexican American: 1 Median :50.00   
:) Mean :0.2667 White :12 Mean :53.33   
:) 3rd Qu.:0.5000 3rd Qu.:65.00   
:) Max. :1.0000 Max. :70.00   
:)   
:) Children SubsUse Alcohol Ment1   
:) Min. :0.0 Min. :0.0000 Min. :1.000 Min. : 7.00   
:) 1st Qu.:0.0 1st Qu.:0.0000 1st Qu.:1.000 1st Qu.:11.00   
:) Median :0.0 Median :0.0000 Median :1.000 Median :14.00   
:) Mean :0.4 Mean :0.2667 Mean :1.067 Mean :13.87   
:) 3rd Qu.:0.5 3rd Qu.:0.5000 3rd Qu.:1.000 3rd Qu.:16.00   
:) Max. :2.0 Max. :1.0000 Max. :2.000 Max. :20.00   
:)   
:) Sport Depr1 Awkward1 Prod2   
:) Min. :0.0000 Min. : 5.00 Min. :1.000 Min. :2.000   
:) 1st Qu.:0.0000 1st Qu.:10.00 1st Qu.:2.000 1st Qu.:3.500   
:) Median :0.0000 Median :12.50 Median :3.000 Median :4.000   
:) Mean :0.2667 Mean :12.29 Mean :2.733 Mean :4.133   
:) 3rd Qu.:0.5000 3rd Qu.:14.75 3rd Qu.:3.500 3rd Qu.:5.000   
:) Max. :1.0000 Max. :19.00 Max. :5.000 Max. :7.000   
:) NA's :1   
:) Ment2 Depr2 Awkward2   
:) Min. :1.000 Min. : 5.00 Min. :0.000   
:) 1st Qu.:4.000 1st Qu.:10.00 1st Qu.:1.000   
:) Median :7.000 Median :12.00 Median :2.000   
:) Mean :6.067 Mean :12.13 Mean :2.067   
:) 3rd Qu.:8.000 3rd Qu.:14.50 3rd Qu.:3.000   
:) Max. :9.000 Max. :18.00 Max. :6.000   
:)

## 5.  
d %>%  
 group\_by(Sex) %>%  
 summarize(Dep = mean(Depr1, na.rm=TRUE),  
 DepSD = sd(Depr1, na.rm=TRUE))

:) # A tibble: 2 x 3  
:) Sex Dep DepSD  
:) <fctr> <dbl> <dbl>  
:) 1 0 12.77778 3.993050  
:) 2 1 11.40000 3.646917

## 6.  
d2 = d %>%  
 select(Name, Sex, Income, Awkward1, Awkward2, Prod1, Prod2, Depr1, Depr2) %>%  
 as.data.frame %>%  
 reshape(varying = list(c("Awkward1", "Awkward2"), c("Prod1", "Prod2"),  
 c("Depr1", "Depr2")),  
 direction = "long",  
 v.names = c("Awkward", "Prod", "Depr"),  
 timevar = "Time",  
 idvar = "Name") %>%  
 arrange(Name, Time)  
  
## 7.   
d2$Name

:) [1] Andy Andy Angela Angela Creed Creed Dwight   
:) [8] Dwight Jim Jim Kelley Kelley Kevin Kevin   
:) [15] Meredith Meredith Michael Michael Oscar Oscar Pam   
:) [22] Pam Phyllis Phyllis Ryan Ryan Stanley Stanley   
:) [29] Toby Toby   
:) 15 Levels: Andy Angela Creed Dwight Jim Kelley Kevin Meredith ... Toby

## 8.   
d2$Sex \* 3

:) [1] NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA NA  
:) [24] NA NA NA NA NA NA NA

## 9.  
## You only want those that had high Depression scores across both waves.  
## How would you subset your data to analyze just those individuals?  
  
d %>%  
 filter(Depr1 > 10 | Depr2 > 10)

:) X Name Prod1 MentalApt PhysApt Married Sex Race Income Children  
:) 1 1 Michael 2 3 8 0 0 White 55 0  
:) 2 4 Dwight 5 6 8 0 0 White 70 0  
:) 3 5 Stanley 4 7 4 1 0 Black 70 1  
:) 4 6 Phyllis 4 8 4 1 1 White 70 0  
:) 5 7 Creed 1 2 4 0 0 White 45 0  
:) 6 8 Meredith 3 5 4 0 1 White 40 1  
:) 7 11 Kevin 2 6 2 0 0 White 45 0  
:) 8 12 Kelley 3 5 5 0 1 Indian 40 0  
:) 9 13 Ryan 2 2 5 0 0 White 40 0  
:) 10 14 Toby 4 1 6 0 0 White 60 0  
:) 11 15 Andy 3 5 7 0 0 White 60 0  
:) SubsUse Alcohol Ment1 Sport Depr1 Awkward1 Prod2 Ment2 Depr2 Awkward2  
:) 1 1 1 10 1 15 5 4 3 14 6  
:) 2 0 1 17 0 12 4 7 7 12 4  
:) 3 0 1 15 0 13 1 5 9 13 0  
:) 4 0 1 15 1 10 1 5 9 11 0  
:) 5 1 1 20 0 NA 3 3 4 11 2  
:) 6 1 2 10 0 13 4 2 6 13 4  
:) 7 0 1 7 0 14 3 2 7 15 2  
:) 8 0 1 10 0 17 2 4 6 17 1  
:) 9 1 1 12 0 11 2 4 2 12 1  
:) 10 0 1 15 0 19 1 5 1 18 0  
:) 11 0 1 13 0 16 3 3 8 16 2

## 10.  
d = d %>%  
 mutate(atrisk = ifelse(Depr1 > 10 & SubsUse == 1, "at risk",  
 ifelse(Depr1 <= 10 | SubsUse == 0, "not at risk", NA)))  
  
furniture::table1(d, Prod1,  
 splitby = ~Sex,  
 type = "condense",  
 output = "markdown")

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
| Observations | 10 | 5 |
| Prod1 | 3.1 (1.4) | 3.4 (0.5) |

d = d %>%  
 filter(Depr1 > 10)

The mean of productivity at wave I was 3.1.

## Section 2

You can also embed plots, for example:

###############  
## Functions ##  
###############  
  
## What are the elements of writing a named function?  
  
## What does the following code do?  
thing = list()  
for (i in 1:10){  
 thing[[i]] = mean(data[, i])  
}  
  
## What about this line of code?  
thing = lapply(data, mean)  
  
  
#############  
## ggplot2 ##  
#############  
  
## Let's practice with ggplot2 :)  
## Take a look at the office data set (long or wide) and think of how you  
## could present the data informatively. Then answer these questions:  
## 1. What type of plot do you want to make?  
## 2. What does the data need to look like to get it ready for this plot  
## (i.e., what format does it need to be in)?  
## 3. What R code would you use to make it?  
  
## What would you expect this plot to look like?  
ggplot(summed\_data, aes(x = time, y = mean,   
 group = gender,   
 fill = gender,  
 color = gender)) +  
 geom\_bar(stat = "identity") +  
 geom\_errorbar(aes(ymin = mean - se, ymax = mean + se))

Note that the echo = FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.