Unit 4 – Introduction to R Studio (Part 2)
Homework
SOLUTIONS

Before you begin: Download from the course website
hersdata.csv

Description of hersdata.csv
These data are a simple random sample of n=2763 observations from the HERS study. The HERS study was a randomized clinical trial of hormone therapy (estrogen plus progestin) for the reduction of cardiovascular disease risk in post-menopausal women with established coronary disease. Study participants were n=2,763 women who were: (1) post-menopausal (2) with coronary disease; and (3) with an intact uterus. *We will be using just a few of the variables in this dataset in this assignment.*

Initialize R Studio Session. Clear the Decks

```
setwd("~/Desktop")                    # Set working directory
getwd()                               # Check working directory
options(scipen=999)                   # Turn off scientific notation
rm(list = ls())                       # Clear the Decks
```

Q1 - Import hersdata.csv

```
library(readr)
hersdata <- read_csv("hersdata.csv")
# str(hersdata)                        # Check.  Not knitted.
```

Q2 - convert drinkany to a factor variable

```
class(hersdata$drinkany)                                          # confirm drinkany is character
## [1] "character"

table(hersdata$drinkany)                                          # inspect
##
##   no  yes
## 1680 1081

hersdata$drinkanyf <- factor(hersdata$drinkany,
                             levels=c("no", "yes"),ordered=TRUE)  # set desired storage


class(hersdata$drinkanyf)                                         # confirm drinkanyf is factor
## [1] "ordered" "factor"

table(hersdata$drinkany,hersdata$drinkanyf)                       # check
##
##         no  yes
##   no  1680    0
##   yes    0 1081
```

Q3 - Using package summarytools, produce a one way frequency of drinkany. How many missings?

```
library(summarytools)
summarytools::freq(hersdata$drinkanyf)

## Frequencies
## hersdata$drinkanyf
## Type: Ordered Factor
##
##               Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- ------ --------- -------------- --------- --------------
##          no   1680     60.85          60.85     60.80          60.80
##         yes   1081     39.15         100.00     39.12          99.93
##        <NA>      2                               0.07         100.00  There are 2 missing values
##       Total   2763    100.00         100.00    100.00         100.00
```

Q4 - Create a numeric variable drinkany01

```
library(gmodels)

hersdata$drinkany01 <- NA
hersdata$drinkany01[hersdata$drinkany=="no"] <- 0
hersdata$drinkany01[hersdata$drinkany=="yes"] <- 1

gmodels::CrossTable(hersdata$drinkanyf,hersdata$drinkany01,digits=2,          # check.
                    prop.r=FALSE,prop.c=FALSE,prop.t=FALSE,prop.chisq=FALSE,
                    dnn=c("drinkany - factor","drinkany01 - numeric"))

##
## Total Observations in Table:  2761
##
##
##                    | drinkany01 - numeric
## drinkany - factor |         0 |         1 | Row Total |
## ------------------|-----------|-----------|-----------|
##                no |      1680 |         0 |      1680 |
## ------------------|-----------|-----------|-----------|
##               yes |         0 |      1081 |      1081 |
## ------------------|-----------|-----------|-----------|
##      Column Total |      1680 |      1081 |      2761 |
## ------------------|-----------|-----------|-----------|
```

Q5 - Label BMI

```
library(Hmisc)
Hmisc::label(hersdata$BMI) <- "Body Mass Index"
```

Q6 - Create a subset of hersdata called mytiny

```
library(tidyverse)

mytiny <- hersdata %>%                          # Start with hersdata, THEN DO
            filter(weight > 125) %>%            # Observations to keep: weight > 125
            select(HT, LDL, SBP)                # Variables to keep:  HT, LDL, SBP
```

Q7 – Print mytiny

```
mytiny
```

```
## # A tibble: 5 x 3
##   HT                LDL   SBP
##   <chr>           <dbl> <dbl>
## 1 hormone therapy  122.   129
## 2 placebo          205.   133
## 3 placebo          161.   112
## 4 placebo          137    130
## 5 placebo          148.   139
```

Q8 - Save to Rdataset names mytiny.Rdata

```
save(mytiny, file="mytiny.Rdata")
```