Week11

Mathi Manavalan

4/6/2020

Libraries

```
library(tidyverse)
```

```
## -- Attaching packages ------------------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.0     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   0.8.4
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0

## -- Conflicts ---------------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:dplyr':
##
##     src, summarize

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

Data Import and Cleaning

First, I am importing the entire dataset.

```
data <- as_tibble(spss.get("../data/GSS2006.sav", use.value.labels=TRUE))
```

Now,we only want the variables relevant to the personality inventory and the respondent's self-reported health. For the personality inventory, I am interpreting this to mean the variables (according to the code book for this dataset) from BIG5A1 to BIG5E2 from *The 2006 Module: Personality Traits*. For the respondent's self-reported health, I am interpreting this to mean the HEALTH variable.

```
clean <- data %>%
  select(BIG5A1, BIG5A2, BIG5B1, BIG5B2, BIG5C1, BIG5C2, BIG5D1, BIG5D2, BIG5E1, BIG5E2, HEALTH) %>%
  na.omit() %>%
  mutate(HEALTH = HEALTH %>%
           factor() %>%
           #fct_rev() %>%
           as.numeric()
         )
```

Responses of 'don't know', 'inapplicable', or other unclearly answered items are appropriately marked as NA according to R, so I removed all rows of data that contain NA (in any column(s)).

All of the variables are in their appropriate data type (factors), with corresponding levels.

Analysis

```
set.seed(1)

# Shuffle row indices: rows
rows <- sample(nrow(clean))

# Randomly order data
shuffled <- clean[rows, ]

holdout <- clean[1:250, ]

train <- clean[251:nrow(clean), ]



olsr <- lm(HEALTH ~ ., holdout)
olsr
```

```
##
## Call:
## lm(formula = HEALTH ~ ., data = holdout)
##
## Coefficients:
##                       (Intercept)                       BIG5A1Agree
##                          2.144651                         -0.070409
## BIG5A1Neither agree nor disagree                    BIG5A1Disagree
##                          0.000958                         -0.043901
##          BIG5A1Strongly disagree                       BIG5A2Agree
##                         -0.224848                          0.173522
```

```
## BIG5A2Neither agree nor disagree                 BIG5A2Disagree
##                          0.400288                     -0.021829
##          BIG5A2Strongly disagree                    BIG5B1Agree
##                          0.362036                     -0.264212
## BIG5B1Neither agree nor disagree                 BIG5B1Disagree
##                         -0.307978                      0.139656
##          BIG5B1Strongly disagree                    BIG5B2Agree
##                         -0.652772                      0.509861
## BIG5B2Neither agree nor disagree                 BIG5B2Disagree
##                          0.469689                      0.421304
##          BIG5B2Strongly disagree                    BIG5C1Agree
##                          0.654539                      0.193824
## BIG5C1Neither agree nor disagree                    BIG5C2Agree
##                          0.731099                     -0.884251
## BIG5C2Neither agree nor disagree                 BIG5C2Disagree
##                         -0.935055                     -0.666508
##          BIG5C2Strongly disagree                    BIG5D1Agree
##                         -0.704198                      0.021551
## BIG5D1Neither agree nor disagree                 BIG5D1Disagree
##                          0.189085                      0.243709
##          BIG5D1Strongly disagree                    BIG5D2Agree
##                          1.321893                      0.401722
## BIG5D2Neither agree nor disagree                 BIG5D2Disagree
##                          0.210145                      0.133441
##          BIG5D2Strongly disagree                    BIG5E1Agree
##                          0.219446                      0.085454
## BIG5E1Neither agree nor disagree                 BIG5E1Disagree
##                         -0.104056                      0.216474
##          BIG5E1Strongly disagree                    BIG5E2Agree
##                         -0.340226                     -0.402309
## BIG5E2Neither agree nor disagree                 BIG5E2Disagree
##                         -0.268693                     -0.338218
##          BIG5E2Strongly disagree
##                         -0.167624
```

Visualization