Clean Complete Cases

Martin Skarzynski

2018-04-19

```r
library(here)
```

```
## here() starts at /Users/marskar/gdrive/nhanes
```

```r
library(readr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

```r
library(caret)
```

```
## Loading required package: lattice

## Loading required package: ggplot2
```

```r
#remove non-numeric variables
#remove variables with more than one unique value
#remove highly correlated variables

dat <- read_rds(here('dat/2-join-complete-cases.rds')) %>%
    select_if(is.numeric) %>%
    select_if(~n_distinct(.) > 1)  %>%
    select(-findCorrelation(cor(.), cutoff=0.82)) %>%
    write_rds(here("dat/3-clean-complete-cases.rds"))
```