

Data Exploration

12/7/2019

```
library(tidyverse)
```

```
## -- Attaching packages ----- t
## v ggplot2 3.1.0      v purrr  0.3.2
## v tibble  2.1.1      v dplyr  0.8.0.1
## v tidyr   0.8.3      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.4.0

## Warning: package 'ggplot2' was built under R version 3.4.4
## Warning: package 'tibble' was built under R version 3.4.4
## Warning: package 'tidyr' was built under R version 3.4.4
## Warning: package 'readr' was built under R version 3.4.4
## Warning: package 'purrr' was built under R version 3.4.4
## Warning: package 'dplyr' was built under R version 3.4.4
## Warning: package 'stringr' was built under R version 3.4.4
## Warning: package 'forcats' was built under R version 3.4.4

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

```
library(magrittr)
```

```
##
## Attaching package: 'magrittr'

## The following object is masked from 'package:purrr':
##
##   set_names

## The following object is masked from 'package:tidyr':
##
##   extract
```

```
library(janitor)
```

```
##
## Attaching package: 'janitor'

## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test
```

```
##data wrangling ----
df <- read_csv("ipums_time_used.csv") %>%
  clean_names() %>%
  filter(!is.na(wb_resp)) %>%
  rename(
    residence = metro,
```

```

education = educ,
employment_status = empstat,
wellbeing_response = wb_resp,
hourly_wage = hourwage,
hours_work_wk = uhrsworkt) %>%
select(-wt06, -pernum, -lineno, -caseid) %>%
mutate(sex = case_when(
  sex == 1 ~ "Male",
  sex == 2 ~ "Female"),
employment_status = case_when(
  employment_status == 1 ~ "Employed",
  employment_status == 2 ~ "Employed",
  employment_status == 3 ~ "Unemployed",
  employment_status == 4 ~ "Unemployed",
  employment_status == 5 ~ "Not in labor force"),
education = case_when(education < 20 ~ "BelowHS",
  education == 20 ~ "High School",
  education == 21 | education == 30 ~ "Some College",
  education == 31 | education == 32 ~ "Associate Degree",
  education == 40 ~ "Bachelor's Degree",
  education == 41 ~ "Master's Degree",
  education == 42 ~ "Professional Degree",
  education == 43 ~ "Doctoral Degree"),
race = case_when(
  race == 100 ~ "White",
  race == 110 ~ "Black",
  race == 120 ~ "American Indian",
  race == 131 ~ "Asian",
  race == 132 ~ "Pacific Islander",
  race == 200 | race == 210 | race == 211 | race == 212 | race == 300 | race == 400 ~ "Black-M",
  race == 201 | race == 202 | race == 203 | race == 310 | race == 320 ~ "White-Mixed",
  race == 220 | race == 230 ~ "Other-Mixed"),
residence = case_when (
  residence == 1 ~ "Metropolitan: Central City",
  residence == 2|residence == 3 ~ "Metropolitan: Others",
  residence == 4 ~ "Nonmetropolitan")) %>%
filter (!is.na(residence))

```

Parsed with column specification:

```

## cols(
##   YEAR = col_double(),
##   CASEID = col_double(),
##   METRO = col_double(),
##   PERNUM = col_double(),
##   LINENO = col_double(),
##   WT06 = col_double(),
##   AGE = col_double(),
##   SEX = col_double(),
##   RACE = col_double(),
##   EDUC = col_double(),
##   EMPSTAT = col_double(),
##   UHRSWORKT = col_double(),
##   HOURWAGE = col_double(),
##   WB_RESP = col_double()

```

```
## )
```

Questions:

Does location effect happiness of a race?

Hypothesis:

Location and Race does not have an effect on someone's happiness $\beta_0 = 0$ $\beta_A \neq 0$

Model 1:

```
race_location <- lm(wellbeing_response ~ race + residence, data = df)
summary(race_location)
```

```
##
## Call:
## lm(formula = wellbeing_response ~ race + residence, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.95186  0.06239  0.06239  0.06517  0.10917
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.930054   0.015242   61.019  <2e-16 ***
## raceAsian      -0.039221   0.016575   -2.366    0.018 *
## raceBlack      -0.013586   0.015463   -0.879    0.380
## raceBlack-Mixed  0.019026   0.025930    0.734    0.463
## raceOther-Mixed  0.068704   0.080872    0.850    0.396
## racePacific Islander -0.017097   0.033612   -0.509    0.611
## raceWhite       0.004777   0.015151    0.315    0.753
## raceWhite-Mixed -0.001122   0.020431   -0.055    0.956
## residenceMetropolitan: Others 0.002783   0.003119    0.892    0.372
## residenceNonmetropolitan  0.001357   0.004135    0.328    0.743
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2513 on 36791 degrees of freedom
## Multiple R-squared:  0.001747,    Adjusted R-squared:  0.001503
## F-statistic: 7.154 on 9 and 36791 DF,  p-value: 1.941e-10
```

Model 1 Analysis:

Model 2:

Model 2 Analysis:

Model 3:

Model 3 Analysis:

Model 4:

Model 4 Analysis: