# Project notepad

**For the project**

David Goh

## Packages

```r
library(tidyverse)
library(tidymodels)
library(knitr)
```

## Load data

```r
abortion_data <- read_csv("data/wvs-usa-abortion-attitudes-data.csv")
glimpse(abortion_data)
```

```
Rows: 10,387
Columns: 16
$ wvsccode         <dbl> 840, 840, 840, 840, 840, 840, 840, 840, 840, 840, 840~
$ wave             <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ year             <dbl> 1982, 1982, 1982, 1982, 1982, 1982, 1982, 1982, 1982,~
$ aj               <dbl> 5, 5, NA, 1, 5, 1, 1, 1, 1, 1, 1, 1, 1, 5, 6, 3, 2, 5~
$ age              <dbl> 40, 43, 18, 18, 22, 21, 37, 45, 30, 72, 22, 47, 56, 5~
$ collegeed        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
$ female           <dbl> 0, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0,~
$ unemployed       <dbl> 0, 0, 0, 0, NA, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0~
$ ideology         <dbl> 8, NA, 10, NA, NA, 4, 8, 5, NA, 2, NA, 7, 10, 10, 10,~
$ satisfinancial   <dbl> 5, 3, 2, 6, 5, 8, 9, 10, 8, 6, 1, 8, 3, 8, NA, 2, 5, ~
$ postma4          <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
$ cai              <dbl> 0, -2, 0, -1, 0, 1, -1, -1, 0, -1, 0, -2, 0, -1, -2, ~
$ trustmostpeople  <dbl> 1, 0, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0,~
$ godimportant     <dbl> 10, 10, 8, 10, 5, 9, 10, 10, 7, 10, 5, 10, 10, 10, 10~
$ respectauthority <dbl> 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
$ nationalpride    <dbl> 1, NA, 1, 1, 1, 0, 0, NA, NA, 1, NA, 0, 1, 1, 1, 1, 1~
```

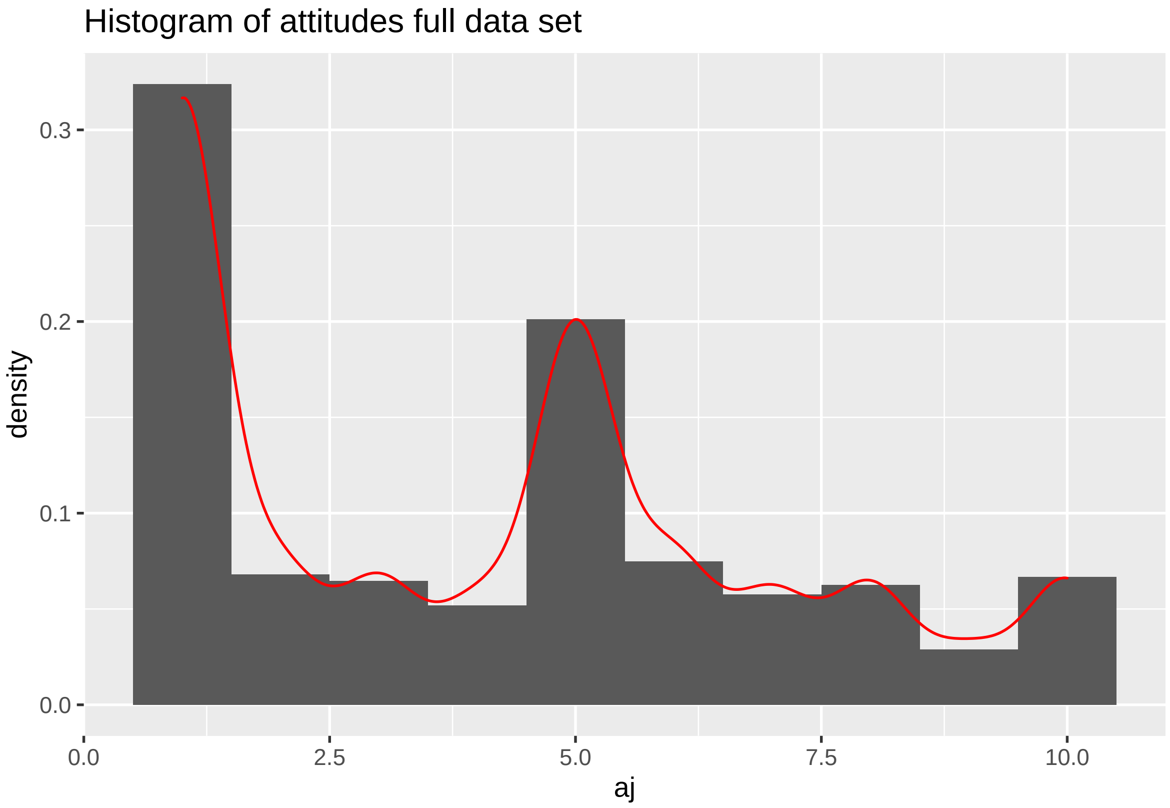
## Exploratory Data Analysis

Visualization and summary statistics for the response variable

```
ggplot(abortion_data, aes(x = aj)) +
  geom_histogram(binwidth = 1, aes(y=..density..)) +
  geom_density(color = "red") +
  labs(title = "Histogram of attitudes full data set")
```

```
Warning: Removed 299 rows containing non-finite values (stat_bin).
```

```
Warning: Removed 299 rows containing non-finite values (stat_density).
```



```
summary(abortion_data$aj)
```

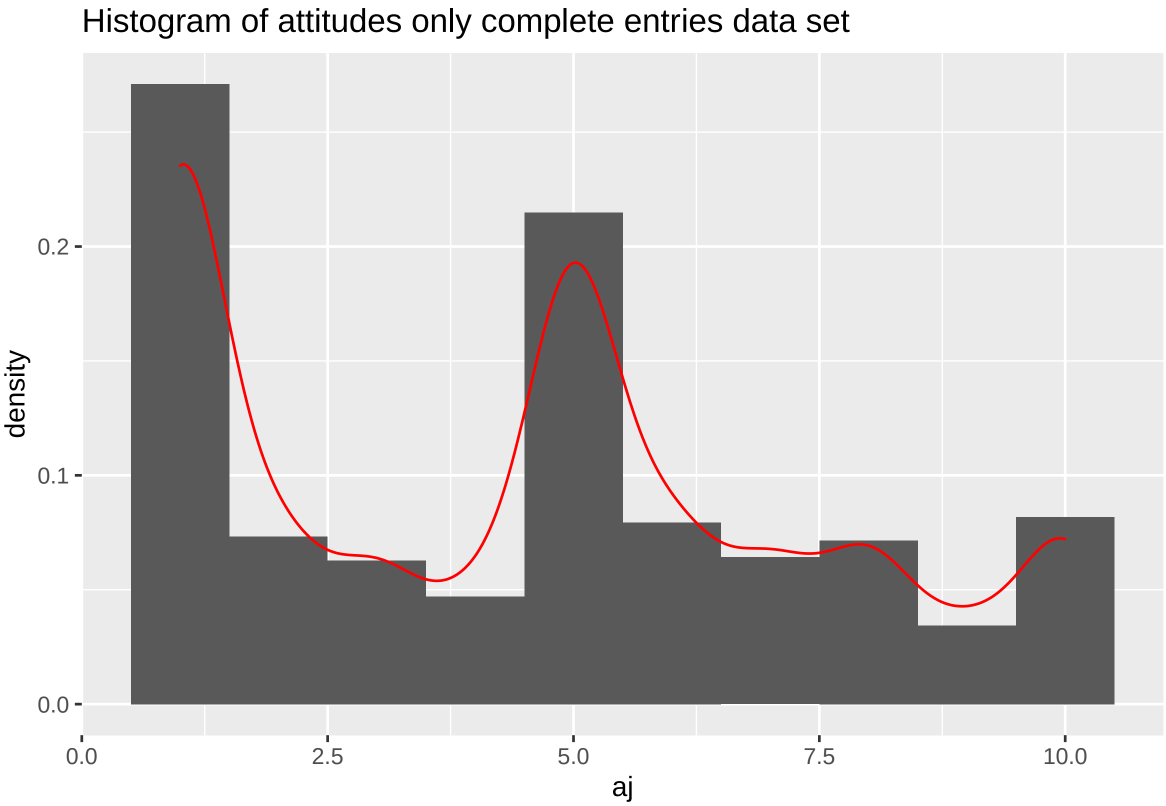
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
  1.000   1.000   4.000   4.147   6.000  10.000     299 
```

### EDA on how to populate "ideology"

We will run all EDA on the set of responses that include ideology values and separately on the set of responses that do not in order to inform how we should populate those values.

```
all_complete <- abortion_data[complete.cases(abortion_data),]

ggplot(all_complete, aes(x = aj)) +
  geom_histogram(binwidth = 1, aes(y=..density..)) +
  geom_density(color = "red") +
  labs(title = "Histogram of attitudes only complete entries data set")
```



```
summary(all_complete$aj)
```

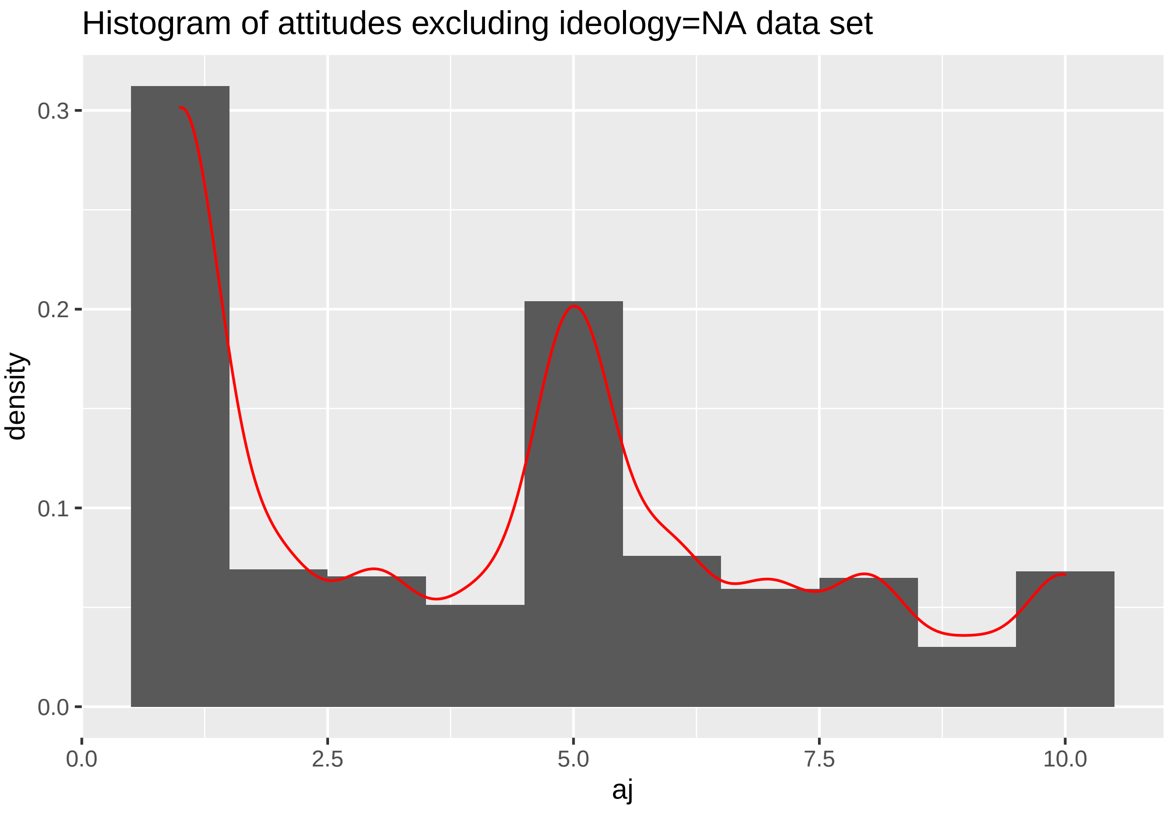
```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  1.000   1.000   5.000   4.492   7.000  10.000 
```

```
no_ideology <- abortion_data[!(abortion_data$ideology==""),]

ggplot(no_ideology, aes(x = aj)) +
  geom_histogram(binwidth = 1, aes(y=..density..))+
  geom_density(color = "red") +
  labs(title = "Histogram of attitudes excluding ideology=NA data set")
```

```
Warning: Removed 1010 rows containing non-finite values (stat_bin).
```

```
Warning: Removed 1010 rows containing non-finite values (stat_density).
```



```
summary(no_ideology$aj)
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
   1.00    1.00    5.00    4.21    6.00   10.00    1010 
```

### List of variables that will be considered as predictors

To discuss as group

### Run MLR on some variables

```
mlr_main_fit <- linear_reg() %>%
  set_engine("lm") %>%
  fit(aj ~ collegeed + female + ideology + trustmostpeople + godimportant, data = abortion_da

tidy(mlr_main_fit)
```

```
# A tibble: 6 x 5
  term            estimate std.error statistic   p.value
  <chr>              <dbl>     <dbl>     <dbl>     <dbl>
1 (Intercept)        8.83     0.144      61.3  0        
2 collegeed          0.666    0.0791      8.42 4.68e- 17
3 female             0.287    0.0700      4.10 4.16e-  5
4 ideology          -0.299    0.0182    -16.5  1.52e- 59
5 trustmostpeople    0.184    0.0717      2.56 1.05e-  2
6 godimportant      -0.372    0.0133    -28.0  3.19e-161
```