# HW2 - Australia Stats Data

John C. Manyiel

9/25/2021

Load packages

```
#library(rprojroot)
library(readxl)
library(rmarkdown)
library(ggplot2)
library(rvest)
library(tibble)
library(lubridate)
library(tidyverse)
library(stringr)
library(pdftools)
library(xlsx)
library(dbplyr)
library(knitr)
theme_set(bayesplot::theme_default(base_family = "sans"))
```

##Read in the data set

```
file_location <- file.path("Z:/Development",
                           "DatumAnalyticConsulting",
                           "DACSS601Fall21",
                           "_data",
                           "australian_marriage_law_postal_survey_2017_-_response_final.xls")
obs_count <-read_excel(
                path = file_location,
                range = "Table 2!A8:P183",
                col_names = c("area",
                              "Yes",
                              "Yes pct",
                              "No",
                                     "No pct",
                                     "Response Total",
                                     "Response Total pct",
                                     "blank",
                                     "Response clear",
                                     "Response clear pst",
                                     "Response not clear(b)",
                                     "Response not clear(b) pct",
                                     "Non-responding",
                                     "Non-responding pct",
```

```
                                    "Eligible Total",
                                    "Eligible Total pct"))


head(obs_count)
# A tibble: 6 x 16
  area      Yes `Yes pct`    No `No pct` `Response Total` `Response Total~ blank
  <chr>   <dbl>     <dbl> <dbl>    <dbl>            <dbl>            <dbl> <lgl>
1 New So~    NA      NA      NA     NA                 NA               NA NA
2 Banks   37736      44.9 46343     55.1            84079              100 NA
3 Barton  37153      43.6 47984     56.4            85137              100 NA
4 Bennel~ 42943      49.8 43215     50.2            86158              100 NA
5 Berowra 48471      54.6 40369     45.4            88840              100 NA
6 Blaxla~ 20406      26.1 57926     73.9            78332              100 NA
# ... with 8 more variables: Response clear <dbl>, Response clear pst <dbl>,
#   Response not clear(b) <dbl>, Response not clear(b) pct <dbl>,
#   Non-responding <dbl>, Non-responding pct <dbl>, Eligible Total <dbl>,
#   Eligible Total pct <dbl>
```

##Transfor Data Set

Subset by Division and add Division column with row value "New South Wales"

```
NewSouthWales<-  obs_count[2:48,] %>% as_tibble() %>%
                                    mutate(Division = "New South Wales")

head(NewSouthWales[,17])
# A tibble: 6 x 1
  Division
  <chr>
1 New South Wales
2 New South Wales
3 New South Wales
4 New South Wales
5 New South Wales
6 New South Wales
```

Subset by Division and add Division column with row value "Victoria"

```
Victoria <-  obs_count[52:88,] %>% as_tibble() %>%
                                    mutate(Division = "Victoria")

head(Victoria[,17])
# A tibble: 6 x 1
  Division
  <chr>
1 Victoria
2 Victoria
3 Victoria
4 Victoria
5 Victoria
6 Victoria
```

Subset by Division and add Division column with row value "Queensland"

```
Queensland <-  obs_count[92:121,] %>% as_tibble() %>%
                                        mutate(Division = "Queensland")

head(Queensland[,17])
# A tibble: 6 x 1
  Division
  <chr>
1 Queensland
2 Queensland
3 Queensland
4 Queensland
5 Queensland
6 Queensland
```

Subset by Division and add Division column with row value “South Australia”

```
SouthAustralia <-  obs_count[125:135,] %>% as_tibble() %>%
                                        mutate(Division = "South Australia")

head(SouthAustralia[,17])
# A tibble: 6 x 1
  Division
  <chr>
1 South Australia
2 South Australia
3 South Australia
4 South Australia
5 South Australia
6 South Australia
```

Subset by Division and add Division column with row value “Western Australia”

```
WesternAustralia <-  obs_count[139:154,] %>% as_tibble() %>%
                                        mutate(Division = "Western Australia")

head(WesternAustralia[,17])
# A tibble: 6 x 1
  Division
  <chr>
1 Western Australia
2 Western Australia
3 Western Australia
4 Western Australia
5 Western Australia
6 Western Australia
```

Subset by Division and add Division column with row value “Tasmania”

```
Tasmania <-  obs_count[158:162,] %>% as_tibble() %>%
                                        mutate(Division = "Tasmania")

head(Tasmania[,17])
```

```
# A tibble: 5 x 1
  Division
  <chr>
1 Tasmania
2 Tasmania
3 Tasmania
4 Tasmania
5 Tasmania
```

Subset by Division and add Division column with row value "Northern Territory"

```
NorthernTerritory <-  obs_count[166:167,] %>% as_tibble() %>%
                                        mutate(Division = "Northern Territory")

head(NorthernTerritory[,17])
# A tibble: 2 x 1
  Division
  <chr>
1 Northern Territory
2 Northern Territory
```

Subset by Division and add Division column with row value "Australian Capital Territory"

```
CapitalTerritory <-  obs_count[171:172,] %>% as_tibble() %>%
                                        mutate(Division = "Australian Capital Territory")


head(CapitalTerritory[,17])
# A tibble: 2 x 1
  Division
  <chr>
1 Australian Capital Territory
2 Australian Capital Territory
```

##Recombine Data Set

Combine all Division variables

```
obs_count01 <-  rbind( NewSouthWales, Victoria, Queensland,
                       SouthAustralia, WesternAustralia,
                       Tasmania, NorthernTerritory,
                       CapitalTerritory) %>%  as_tibble()


dim(obs_count01)
[1] 150  17
```

Continue to transform the data set and remove blank field such as total and the blank value

```
obs_count02 <-  obs_count01 %>%
                    select(-blank)%>%
                          filter(!grepl("Total", area))%>%
```

```
                                filter(!grepl("Total", area)) %>%drop_na()

head(obs_count02)
# A tibble: 6 x 16
  area        Yes `Yes pct`    No `No pct` `Response Total` `Response Total pct`
  <chr>     <dbl>     <dbl> <dbl>    <dbl>            <dbl>                <dbl>
1 Banks     37736      44.9 46343     55.1            84079                  100
2 Barton    37153      43.6 47984     56.4            85137                  100
3 Bennelong 42943      49.8 43215     50.2            86158                  100
4 Berowra   48471      54.6 40369     45.4            88840                  100
5 Blaxland  20406      26.1 57926     73.9            78332                  100
6 Bradfield 53681      60.6 34927     39.4            88608                  100
# ... with 9 more variables: Response clear <dbl>, Response clear pst <dbl>,
#   Response not clear(b) <dbl>, Response not clear(b) pct <dbl>,
#   Non-responding <dbl>, Non-responding pct <dbl>, Eligible Total <dbl>,
#   Eligible Total pct <dbl>, Division <chr>
dim(obs_count02)
[1] 150  16
```

#Visualizing Data Set

Begin summarizing & visualizing your data