R

**Kate’s code Percentage for Yes**

library(dplyr)

response <- read.csv("AMLPS\_RESP\_2017\_17112017134913689.csv", header=TRUE, stringsAsFactors = FALSE)

statefilter = c("New South Wales","Victoria","Queensland","South Australia","Western Australia", "Northern Territory" ,"Tasmania", "Australian Capital Territory" )

response\_by\_state <- response %>%

filter(Federal.Electoral.Division %in% statefilter)

yes\_by\_state <- response\_by\_state[response\_by\_state$ï..RESPONSE\_CAT== "RESPCLR\_Y" & response\_by\_state$Measure== "Percentage (%)", c(8,11)]

#make a plot

yes\_by\_state%>%

ggplot()+

geom\_bar(aes(x= Federal.Electoral.Division,y= Value , fill=Federal.Electoral.Division), stat = "identity") +

xlab("State") + ylab("Yes (%)") +

ggtitle("Yes responses Per State")+

theme\_classic()+

theme(axis.text.x = element\_text(angle = 45, hjust = 1))+

coord\_cartesian ()+

theme( plot.title = element\_text(hjust=0.5) )+

theme(panel.border = element\_blank(), panel.grid.major = element\_blank(), panel.grid.minor = element\_blank(), axis.line = element\_line(colour = "grey"))+

theme(panel.border = element\_blank(), axis.line.y = element\_line(colour = "grey"))

**Sadia’s code for Percentage of yes/no per state**

ibrary(dplyr)

states <- c('Victoria (Total)','New South Wales (Total)', 'Queensland (Total)',

'South Australia (Total)', 'Western Australia (Total)', 'Tasmania (Total)',

'Northern Territory (Total)', 'Australian Capital Territory (Total)')

ResponseFemale$Gender <- 'F'

ResponseMale$Gender <- 'M'

ResponseFemale2 <- select(ResponseFemale, -18)

ResponseMale2 <- select(ResponseMale, -18)

ResponseFM <- rbind(ResponseFemale2, ResponseMale2)

StatesGender <- ResponseFM %>%

group\_by(FederalElectoralDivison) %>%

filter(FederalElectoralDivison %in% states)

StatesGender2 <- StatesGender %>% distinct()

**Liam’s codes for map**

## install.packages("")

library(dplyr)

library(rgdal)

library(tmap)

#---------------------------------------------------------------------------#

### Call data

total <- read.csv("ResponseTotal.csv")

map <- readOGR(dsn="101\_electoral\_boundaries/COM20111216\_ELB\_region.shp",layer="COM20111216\_ELB\_region")

# lots of duplicates in the data so we first need to get our data clean

dup\_free\_female = distinct(female)

dup\_free\_male = distinct(male)

dup\_free\_total = distinct(total)

# Find out what variables are measures

unique(total$Participation)

# Filter on different measurement parametes

total\_sub <- dup\_free\_total[which(dup\_free\_total$Participation == "Participation rate (%)"),]

# Make column names the same to merge

total$ELECT\_DIV <- total$FederalElectoralDivison

# Merge data

map1 <- merge(map, total\_sub, by = "ELECT\_DIV")

# Filter on state

map\_vic <- map1[which(map1$STATE == "VIC"),]

# find variables

names(map1)

# Create map using 'fill' to define the variable

qtm(shp = map\_vic, fill = "X18.19.years", fill.palette = "-Blues")

qtm(shp = map\_vic, fill = c("X18.19.years", "X75.79.years"), fill.palette = "Blues", ncol = 2)

**Kate’s code to run a regression of gender ratio by Yes vote and then plot it on a scatterplot**

#some code, based on Sadia's code and Kate's to merge two dataframe's to combine info about   
#total male and female responses, state and yes vote  
  
####First: Sadia's modified code to get a dataframe with state, and male response numbers and female response numbers  
  
#read in  
response\_female <- read.csv("MarriageEqualityResponseFemale.csv", header=TRUE, stringsAsFactors = FALSE)  
response\_male <- read.csv("MarriageEqualityResponseMale.csv", header=TRUE, stringsAsFactors = FALSE)  
  
# duplicate removal  
dup\_free\_female = distinct(response\_female)  
dup\_free\_male = distinct(response\_male)  
  
  
#make a state filter  
states = c("New South Wales (Total)", "Victoria (Total)", "Queensland (Total)", "South Australia (Total)",   
 "Western Australia (Total)", "Tasmania (Total)", "Northern Territory (Total)",   
 "Australian Capital Territory (Total)", "Australia Total")  
# data selection  
test\_m = dup\_free\_male %>%   
 filter(!FederalElectoralDivison %in% states, Participation %in% "Total participants")  
test\_fm = dup\_free\_female %>%  
 filter(!FederalElectoralDivison %in% states, Participation %in% "Total participants")  
# get % participation response  
a\_m = test\_m %>%  
 filter(Participation == "Total participants")   
a\_fm = test\_fm %>%  
 filter(Participation == "Total participants")  
  
for\_plot = cbind(a\_m$FederalElectoralDivison,a\_m$Total.Males.b.,a\_fm$Total.Females.b.)  
for\_plot = as.data.frame(for\_plot)   
for\_plot$V2 = as.numeric(as.character(for\_plot$V2))  
for\_plot$V3 = as.numeric(as.character(for\_plot$V3))  
  
#Rename columns to match my other dataframe  
colnames(for\_plot) <- c("Electorate","Total\_Male\_Response", "Total\_Female\_Reponse")  
  
##Why can't get rid of Australia??  
  
##calculate the female to male participation ratio  
  
for\_plot$Female\_Ratio <- for\_plot[,3]/for\_plot[,2]  
  
  
##Add in State column  
  
for\_plot$State <-NA   
for\_plot$State[1:47] <- "New South Wales"  
for\_plot$State[48:84] <- "Victoria"  
for\_plot$State[85:114] <- "Queensland"  
for\_plot$State[115:125] <- "South Australia"  
for\_plot$State[126:141] <- "Western Australia"  
for\_plot$State[142:146] <- "Tasmania"  
for\_plot$State[147:148] <- "Northern Territory"  
for\_plot$State[149:150] <- "Australian Capital Territory"  
  
  
####Second: Kate's modified code to get Yes vote  
  
#read in  
response <- read.csv("AMLPS\_RESP\_2017\_17112017134913689.csv", header=TRUE, stringsAsFactors = FALSE)  
  
#filter out the state (!)  
statefilter = c("New South Wales","Victoria","Queensland","South Australia","Western Australia",  
 "Northern Territory" ,"Tasmania", "Australian Capital Territory" ,"Australia")  
response\_by\_electorate <- response %>%  
 filter(!Federal.Electoral.Division %in% statefilter)  
  
#create new dataframe

yes\_by\_electorate <- response\_by\_electorate[response\_by\_electorate$ï..RESPONSE\_CAT== "RESPCLR\_Y" &   
 response\_by\_electorate$Measure== "Percentage (%)", c(8,11)]  
  
#Rename   
#Rename columns to match the other dataframe  
colnames(yes\_by\_electorate) <- c("Electorate","Yes")  
  
#count how many rows check the same before we merge (but doesn’t actually matter I found out it still merges)  
  
nrow(yes\_by\_electorate)  
nrow(for\_plot)  
  
  
# merge data frames by Electorate  
gender\_yes\_electorate <- merge(for\_plot, yes\_by\_electorate, by="Electorate")  
nrow(gender\_yes\_electorate)  
#Lost some rows...why?   
  
  
#Run a linear regression  
summary(lm(Yes~Female\_Ratio, data=gender\_yes\_electorate ))

#Plot result  
ggplot(gender\_yes\_electorate, aes(x=Female\_Ratio, y=Yes, label=Electorate )) +  
 geom\_point(aes(colour=State)) +   
 geom\_smooth(method=lm) +  
 theme\_classic()+  
 xlab("Ratio of Females to Males") + ylab("Yes (%)") +  
 ggtitle("Does the Ratio of Female to Male Voters Relate to the Outcome of the Marriage Equality Vote?")+  
 geom\_text(  
 data= gender\_yes\_electorate[gender\_yes\_electorate$Yes <40 | gender\_yes\_electorate$Yes >75,],  
 size= 4, hjust=-0.1,vjust=0.5, check\_overlap = TRUE, aes(colour = factor(State)))