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

#Steps and corresponding actions#

#1. Remove the duplicates in dataset#

#2A. Percentage of Yes per state#

#2B. Bar chart for above percentage#

#3. Per state participation#

#4. Mapping#

#5. Relationship between gender ratio & Yes rate#

-------------------------------------------------------------------------------------------------

#1. Remove the duplicates#

library(dplyr)

female = read.csv("C:/Users/Administrator/Desktop/Marriage Equality/ResponseFemale.csv")

male = read.csv("C:/Users/Administrator/Desktop/Marriage Equality/ResponseMale.csv")

total = read.csv("C:/Users/Administrator/Desktop/Marriage Equality/ResponseTotal.csv")

dup\_free\_female = distinct(female)

dup\_free\_male = distinct(male)

dup\_free\_total = distinct(total)

#2A.Yes per state#

library(dplyr)

AMLPS <- read.csv("C:/Users/Administrator/Desktop/Marriage Equality/AMLPS\_RESP\_2017\_17112017134913689.csv", header=TRUE, stringsAsFactors = FALSE)

state = c("New South Wales", "Victoria", "Queensland", "South Australia",

"Western Australia", "Tasmania", "Northern Territory",

"Australian Capital Territory", "Australia")

test\_ampls = AMLPS %>%

filter(Response %in% c("Response clear - Total", "Response clear - Yes", "Response clear - No")) %>%

filter(Measure == 'Percentage (%)' & `Federal.Electoral.Division` %in% state)

yes <- test\_ampls[which(test\_ampls$Response == "Response clear - Yes"),]

no <- test\_ampls[which(test\_ampls$Response == "Response clear - No"),]

-----------------------------------------------------------------------------------------------------

#2B Bar chart#

library(ggplot2)

#Yes#

yes %>%

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))+

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"))+

scale\_fill\_discrete(name = "State")

#No#

no %>%

ggplot()+

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

xlab("State") + ylab("No (%)") +

ggtitle("No Responses Per State")+

theme\_classic()+

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

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")) +

scale\_fill\_discrete(name = "State")

#Ideas for changing plot

#Order from highest to lowest plot

-----------------------------------------------------------------------------------------------------------

#3. Per state participation#

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% "Participation rate (%)")

test\_fm = dup\_free\_female %>%

filter(FederalElectoralDivison %in% states, Participation %in% "Participation rate (%)")

# get % participation response

a\_m = test\_m %>%

filter(Participation == "Participation rate (%)")

a\_fm = test\_fm %>%

filter(Participation == "Participation rate (%)")

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))

install.packages("tidyr")

library(tidyr)

for\_plot %>%

gather(V2,V3,key='gender',value="participation (%)") %>%

ggplot() +

geom\_bar(aes(y=`participation (%)`,x=V1, fill=gender),stat = 'identity',position = "dodge")+

coord\_flip()+

scale\_fill\_manual(values = c("blue", "Red"), name = "State/Gender", labels = c("Male ","Female"))

#is not working :( need to check again#

------------------------------------------------------------------------------------------------------

# install.packages("")

install.packages("rgdal")

install.packages("tmap")

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")

# 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)

--------------------------------------------------------------------------------------------------------

#5. Relationship between gender ratio & Yes rate#

#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

#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] <- "New South Wales"

for\_plot$State[2] <- "Victoria"

for\_plot$State[3] <- "Queensland"

for\_plot$State[4] <- "South Australia"

for\_plot$State[5] <- "Western Australia"

for\_plot$State[6] <- "Tasmania"

for\_plot$State[7] <- "Northern Territory"

for\_plot$State[8] <- "Australian Capital Territory"

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*????#Need to check the following commands#\*\*

#read in

response <- read.csv("C:/Users/Administrator/Desktop/Marriage Equality/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$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)))