Characteristics of Early + Mid # of Gos

Theresa Cheng

August 23, 2016

# This script loads, cleans, and plots data related to the number of Gos in early and mid yellow light game behavior ("EM"). It examines relationships between EM, age, sex, and RPI in preparation for a Flux 2016 conference poster.

# Load packages

library(tidyr)  
library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(ggplot2)  
library(sjPlot)

# Acquire data

rm(list=ls())  
  
# Cyberball subject list  
setwd("/Volumes/research/tds/behavioral/raw/cyberball")  
subjectList=as.data.frame(read.csv("SubjectList\_full.csv", header=FALSE), header=FALSE)  
colnames(subjectList) = c("SID", "Age")  
  
# Demographic data  
setwd("/Users/theresacheng/Dropbox (PfeiBer Lab)/TDS/TDS-II/data/Qualtrics/Scored\_Qualtrics/Session\_2")  
df\_demo<-as.data.frame(read.csv("TDS2\_S2\_Qualtrics\_2015.10.03.csv", header=TRUE))  
df\_demo<-df\_demo[,c("SID","Age","Gender")]  
   
# Trial-by-trial stoplight behavior  
setwd("/Volumes/research/tds/behavioral/processed/ylg")  
df\_trial<-read.csv("tds2-all\_trial\_by\_trial.csv", header=TRUE)  
  
# Calculate the number of decisions per subject per run by intersection type  
df\_trial$decision=as.numeric(df\_trial$decision) #make it into a numeric value; 3=stop, 1=go, 2= penalty   
numGo=summarise(group\_by(df\_trial, subject.name, run\_index, type), sum(decision=="1"))   
colnames(numGo)=c("SID", "run","type", "numGo")  
  
# Dealing with NAs, which make it so that the spread function doesn't work  
 sum(is.na(numGo$type))

## [1] 1

numGo=numGo[-1408,]  
  
numGo= spread(numGo, key=type, value=numGo)  
  
# RPI data  
setwd("/Volumes/research/tds/behavioral/processed/cyberball")  
RPI<-read.csv("RPI.csv", header=TRUE)

# Manipulate data for this particular analysis

# Subset demographic information by Cyberball subject list  
df\_demo=merge(subjectList, df\_demo, by="SID")  
df\_demo$Age.y=NULL  
colnames(df\_demo)=c("SID", "age", "sex")  
  
# Create a dataframe of Gos fo early + mid trials only  
numGo$EM=numGo$E+numGo$M # merge the count for early and mid trials  
df=cbind(numGo[,1:2], numGo[,6]) # create a dataframe of just these trials  
df=spread(df, key=run, value=EM)  
  
# Merge with demographic data  
df=merge(df\_demo, df)  
  
# Create a long version for graphing funsies  
df\_long=gather(df, "run", "EM", 4:11)  
  
# Generate variables of average num of Gos for each condition  
df$aloneGo=(df$"3"+df$"4")/2  
df$peerGo=(df$"5"+df$"6")/2  
df$excGo=(df$"7"+df$"8")/2  
  
# Missing data:   
# Note that 189 was removed for lacking ylg beh data  
# 127 and 168 is missing run 8, use run 7 data as the excGo value  
df[10,]$excGo=df[10,]$"7"  
df[43,]$excGo=df[43,]$"7"  
  
# subset df down to grouped variables  
df[,4:11]=NULL  
  
# generate difference scores  
df$peer\_alone=df$peerGo-df$aloneGo  
df$exc\_peer=df$excGo-df$peerGo  
df$exc\_alone=df$excGo-df$aloneGo  
  
# remove 189 from RPI list  
RPI=RPI[-61,]  
  
# add RPI variable to the df  
df$RPI=RPI$RPI\_cov

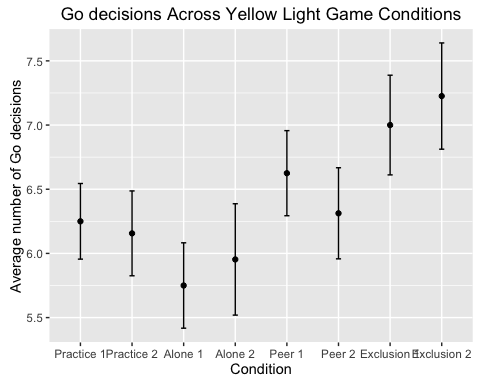
# How do the number of Gos in early and middle intersections change by run?

# Generate a summary dataframe  
EMbyRun=summarise(group\_by(df\_long, run), mean(EM, na.rm=TRUE), se=sd(EM, na.rm=TRUE)/sqrt(length(EM)))  
colnames(EMbyRun)=c("run", "mean", "se")  
  
factor(EMbyRun$run, labels=c("Alone 1", "Alone 2", "Alone 3", "Alone 4", "Peer 1", "Peer 2", "Exclusion 1", "Exclusion 2"))

## [1] Alone 1 Alone 2 Alone 3 Alone 4 Peer 1 Peer 2   
## [7] Exclusion 1 Exclusion 2  
## 8 Levels: Alone 1 Alone 2 Alone 3 Alone 4 Peer 1 Peer 2 ... Exclusion 2

p=ggplot(data=EMbyRun, aes(x=run, y=mean)) + #plot it  
 geom\_errorbar(aes(ymin=mean-se, ymax=mean+se), width=.1) +  
 geom\_point() +  
 geom\_line() +  
 ggtitle("Go decisions Across Yellow Light Game Conditions") +   
 labs(x= "Condition", y="Average number of Go decisions")  
  
p + scale\_x\_discrete(labels=c("Practice 1", "Practice 2", "Alone 1", "Alone 2", "Peer 1", "Peer 2", "Exclusion 1", "Exclusion 2"))

## geom\_path: Each group consists of only one observation. Do you need to  
## adjust the group aesthetic?

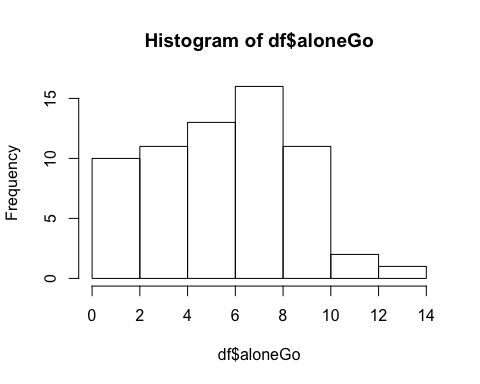


# This plot suggests that we should not lump across the alone trials (1-4), and that differences related to learning the task and/or the scanner environment make 3-4 qualitatively different.   
# Also, this graph suggests that collapsing across 34, 56, and 78 trials is reasonable

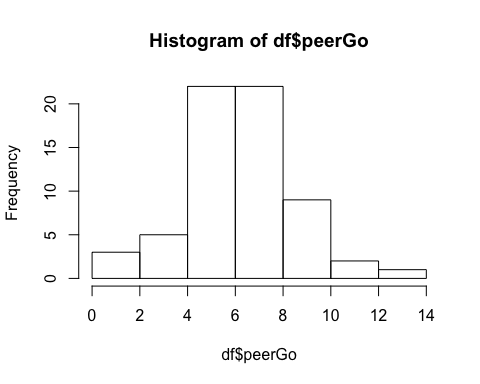
# How is ylg behavior in early+mid #Gos affected by age and sex?

## Descriptives

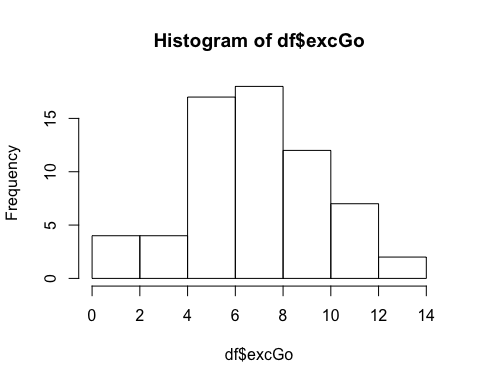
# Inspect the data distributions  
hist(df$aloneGo) #ylg beh



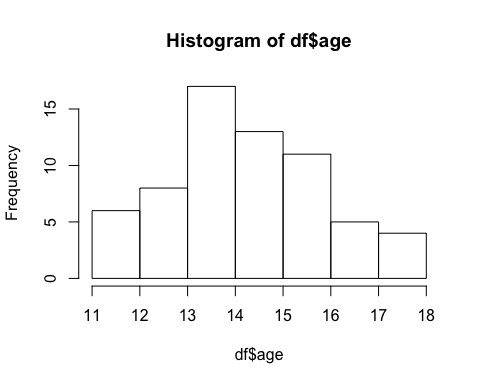
hist(df$peerGo)



hist(df$excGo)



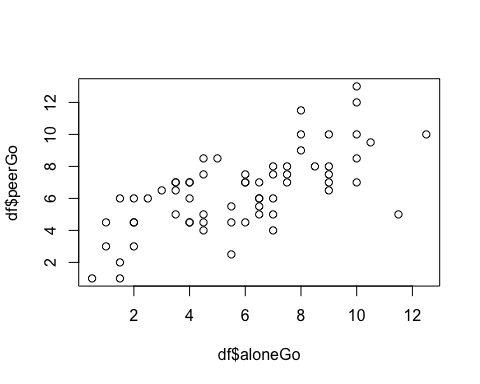
hist(df$age) # age and gender



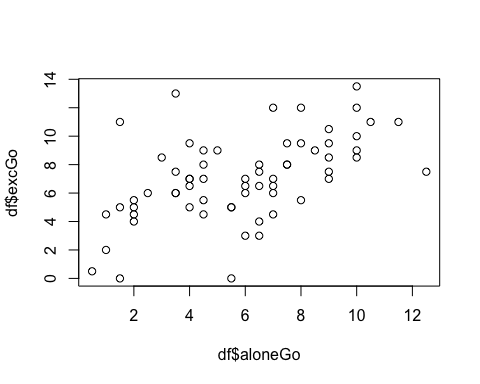
tally(df, sex)

## n  
## 1 31

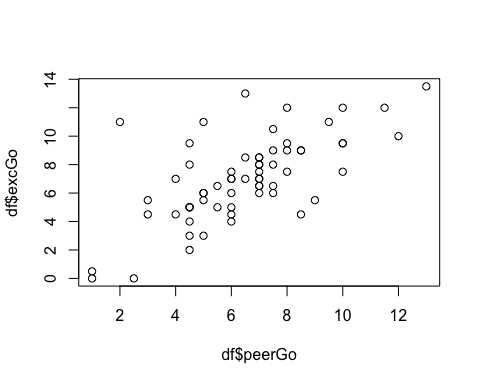
# Inspect scatterplots   
plot(df$aloneGo, df$peerGo) #ylg beh



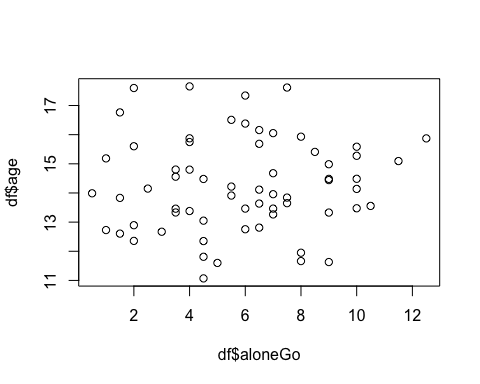
plot(df$aloneGo, df$excGo)



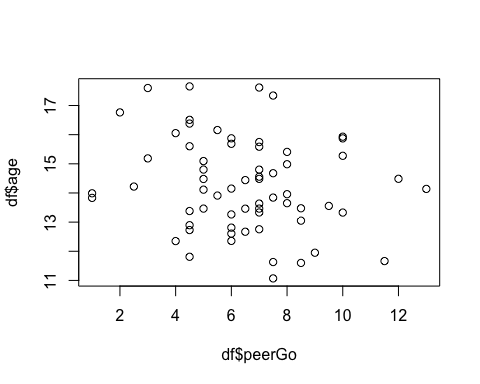
plot(df$peerGo, df$excGo)



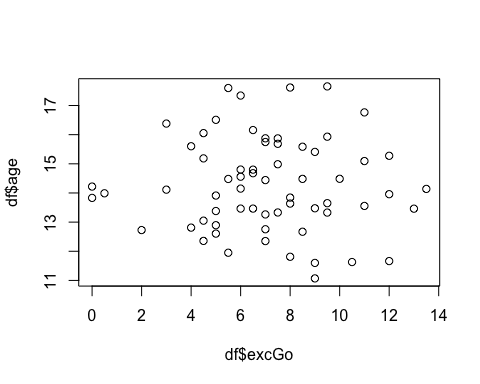
plot(df$aloneGo, df$age) #ylg and age



plot(df$peerGo, df$age)



plot(df$excGo, df$age)



# Examine correlations  
cor.table=cor(df)  
sjt.corr(df, p.numeric=TRUE, corr.method="pearson")

## Regression models

# Do age and sex significantly predict the number of Gos (examined by condition)?  
M1 <- lm(aloneGo ~ age + sex, data=df) #alone  
summary(M1)

##   
## Call:  
## lm(formula = aloneGo ~ age + sex, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.2948 -2.1216 0.1941 2.0810 6.8122   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 4.4819 3.3317 1.345 0.184  
## age 0.1194 0.2312 0.516 0.608  
## sex -0.6889 0.7319 -0.941 0.350  
##   
## Residual standard error: 2.926 on 61 degrees of freedom  
## Multiple R-squared: 0.01826, Adjusted R-squared: -0.01393   
## F-statistic: 0.5673 on 2 and 61 DF, p-value: 0.57

M2 <- lm(peerGo ~ age + sex, data=df) #peer  
summary(M2)

##   
## Call:  
## lm(formula = peerGo ~ age + sex, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.8289 -1.4530 0.0357 1.5057 6.2591   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 10.7776 2.7428 3.929 0.00022 \*\*\*  
## age -0.2855 0.1904 -1.500 0.13881   
## sex -0.4856 0.6025 -0.806 0.42344   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.408 on 61 degrees of freedom  
## Multiple R-squared: 0.04606, Adjusted R-squared: 0.01478   
## F-statistic: 1.473 on 2 and 61 DF, p-value: 0.2374

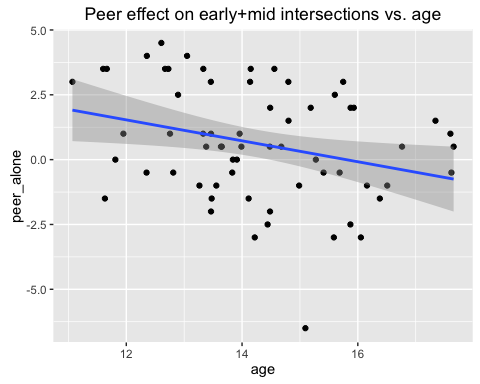
M3 <- lm(excGo ~ age + sex, data=df) #exc  
summary(M3)

##   
## Call:  
## lm(formula = excGo ~ age + sex, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7.2815 -1.8988 0.1085 1.7551 6.2365   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.09001 3.40301 2.377 0.0206 \*  
## age -0.05846 0.23617 -0.248 0.8053   
## sex -0.41541 0.74756 -0.556 0.5805   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.988 on 61 degrees of freedom  
## Multiple R-squared: 0.006115, Adjusted R-squared: -0.02647   
## F-statistic: 0.1877 on 2 and 61 DF, p-value: 0.8294

# Do age and sex significantly predict the differences in the number of Gos (examined by condition)?  
M4 <- lm(peer\_alone ~ age + sex, data=df) #peer minus alone  
summary(M4)

##   
## Call:  
## lm(formula = peer\_alone ~ age + sex, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.6842 -1.4543 -0.0794 1.8718 3.1048   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.2956 2.4316 2.589 0.0120 \*  
## age -0.4049 0.1688 -2.399 0.0195 \*  
## sex 0.2033 0.5342 0.381 0.7048   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.135 on 61 degrees of freedom  
## Multiple R-squared: 0.08777, Adjusted R-squared: 0.05786   
## F-statistic: 2.934 on 2 and 61 DF, p-value: 0.0607

# Yes, graph this effect  
ggplot(data=df, aes(x=age, y=peer\_alone))+  
 geom\_point()+  
 geom\_smooth(method="lm") +  
 ggtitle("Peer effect on early+mid intersections vs. age")



M5 <- lm(exc\_peer ~ age + sex, data=df) #exc minus peer  
summary(M5)

##   
## Call:  
## lm(formula = exc\_peer ~ age + sex, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.3455 -1.2897 -0.0369 0.6920 7.8112   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -2.68759 2.57925 -1.042 0.302  
## age 0.22705 0.17900 1.268 0.209  
## sex 0.07016 0.56660 0.124 0.902  
##   
## Residual standard error: 2.265 on 61 degrees of freedom  
## Multiple R-squared: 0.02603, Adjusted R-squared: -0.005899   
## F-statistic: 0.8153 on 2 and 61 DF, p-value: 0.4473

M6 <- lm(exc\_alone ~ age + sex, data=df) #exc minus alone  
summary(M6)

##   
## Call:  
## lm(formula = exc\_alone ~ age + sex, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.5793 -1.5136 -0.2386 1.7542 8.5997   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 3.6081 3.2659 1.105 0.274  
## age -0.1778 0.2267 -0.785 0.436  
## sex 0.2735 0.7174 0.381 0.704  
##   
## Residual standard error: 2.868 on 61 degrees of freedom  
## Multiple R-squared: 0.01215, Adjusted R-squared: -0.02024   
## F-statistic: 0.375 on 2 and 61 DF, p-value: 0.6889

# Acquire difference scores between conditions in early and mid trials, plus peer minus alone scores with an age residual

df$mc\_age=scale(df$age, scale=FALSE)  
df$mc\_aloneGo=scale(df$aloneGo, scale=FALSE)  
df$mcPeer\_alone=scale(df$peer\_alone, scale=FALSE)  
df$mcExc\_peer=scale(df$exc\_peer, scale=FALSE)  
df$mcExc\_alone=scale(df$exc\_alone, scale=FALSE)  
  
setwd("/Volumes/research/tds/behavioral/processed/cyberball")  
#write.csv(df, "ylgBeh\_earlyMid.csv")  
  
M7 <- lm(peer\_alone ~ age, data=df) #peer minus alone  
summary(M7)

##   
## Call:  
## lm(formula = peer\_alone ~ age, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.7836 -1.3978 -0.1211 1.8815 3.2116   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.3770 2.4054 2.651 0.0102 \*  
## age -0.4037 0.1676 -2.409 0.0190 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.12 on 62 degrees of freedom  
## Multiple R-squared: 0.0856, Adjusted R-squared: 0.07085   
## F-statistic: 5.804 on 1 and 62 DF, p-value: 0.01897

peer\_alone\_ageResid=as.data.frame(as.numeric(M7$residual))  
colnames(peer\_alone\_ageResid)="peer\_alone\_ageResid"  
peer\_alone\_ageResid$mcPeer\_alone\_ageResid=scale(peer\_alone\_ageResid, scale=FALSE)  
  
#write.csv(peer\_alone\_ageResid, "peer\_alone\_ageResid.csv")

# Is the difference between Gos in the exclusion and peer conditions collinear with alone Gos?

cor.test(df$exc\_peer, df$aloneGo)

##   
## Pearson's product-moment correlation  
##   
## data: df$exc\_peer and df$aloneGo  
## t = -0.24649, df = 62, p-value = 0.8061  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.2749832 0.2161832  
## sample estimates:  
## cor   
## -0.03128879

# Now let's think about RPI

summary(M4) #peer minus alone by age + sex, presented for comparison

##   
## Call:  
## lm(formula = peer\_alone ~ age + sex, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.6842 -1.4543 -0.0794 1.8718 3.1048   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.2956 2.4316 2.589 0.0120 \*  
## age -0.4049 0.1688 -2.399 0.0195 \*  
## sex 0.2033 0.5342 0.381 0.7048   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.135 on 61 degrees of freedom  
## Multiple R-squared: 0.08777, Adjusted R-squared: 0.05786   
## F-statistic: 2.934 on 2 and 61 DF, p-value: 0.0607

M5 <- lm(peer\_alone ~ age + RPI + sex, data=df) #peer minus alone  
summary(M5)

##   
## Call:  
## lm(formula = peer\_alone ~ age + RPI + sex, data = df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -6.3828 -1.4860 -0.0255 1.7748 3.5467   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.1074 2.4423 2.501 0.0151 \*  
## age -0.3912 0.1696 -2.307 0.0245 \*  
## RPI -0.5344 0.5704 -0.937 0.3526   
## sex 0.2036 0.5347 0.381 0.7048   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.137 on 60 degrees of freedom  
## Multiple R-squared: 0.1009, Adjusted R-squared: 0.05597   
## F-statistic: 2.245 on 3 and 60 DF, p-value: 0.09225

df\_ofInterest= as.data.frame(cbind(df$age, df$sex, df$RPI, df$peer\_alone))  
colnames(df\_ofInterest)= c("Age", "Sex","RPI","Peer Effect")  
  
cor.table2=cor(df\_ofInterest)  
cor.table2

## Age Sex RPI Peer Effect  
## Age 1.00000000 0.01867468 0.08598479 -0.29257791  
## Sex 0.01867468 1.00000000 0.00211623 0.04107241  
## RPI 0.08598479 0.00211623 1.00000000 -0.13938693  
## Peer Effect -0.29257791 0.04107241 -0.13938693 1.00000000