A Gender Comparison of College Student Concussion Knowledge Among All Athletic Participation Levels

Miami University, Department of Statistics

311 Upham Hall

100 Bishop Circle

Oxford, OH 45056

Wave 2, Team 6:
Tiantian Qiu
James Dudley
Jiahao Sun
Stephen Korwin

STA475 Spring 2018

April 26th, 2018

Contents

1. Abstract

The purpose of this study is to evaluate the current-level of brain injury knowledge among college students and student athletes divided by different genders regarding the definitions, symptoms, recovery patterns and support services available to individuals following injury and to further define the concussion history in this population. The data was collected by through a Qualtrics survey and was administered to collegiates at Miami University and the University of Minnesota. Data visualization, chi-square test for independence, two-way ANOVA, and paired t-test are used in order to compare and show the differences between two genders of people and provide findings efficiently. Aside from one particular question in the survey, there is no significant difference in general concussion knowledge between males and females. Additionally, there is no evidence to suggest that the number of concussions reported by survey respondents changed from beginning to end of the survey.

2. Introduction

This dataset contains the results of a concussion knowledge research study administered through the Qualtrics questionnaire system to varsity athletes, recreational athletes, and non-athletes at Miami University and the University of Minnesota. The purpose of this study is to evaluate the current-level of brain injury knowledge among college students and student athletes regarding the signs, symptoms, recovery patterns, and support services available to individuals following injury and to further define the concussion history in this population. A secondary aim of the study was to determine which educational formats are preferred by collegiate respondent groups when receiving information and education on concussions.

The respondents involved in the survey come from a variety of backgrounds. Of the 306 collegiate students that completed the survey, 206 (67.3%) were female, 99 (32.4%) were male, and 1 (0.3%) did not report. The mean age of the participants was 20.08 years (SD = 2.74 years) and the educational levels consisted of 63 freshman (20.6%), 92 sophomores (30.1%), 62 juniors (20.3%), 68 seniors (22.2%), and 14 graduate students (4.6%). There were 60 (19.6%) varsity athletes representing 12 varsity sports, 190 (62.1%) recreational athletes representing 30 recreational sports, and 56 (18.3%) non-athletes. Respondents from Miami University outnumbered those from the University of Minnesota 297 to 9.

The survey itself consisted of various unique sections: demographic information, self-perception of concussion knowledge, concussion definition knowledge, concussion symptom knowledge, related professionals knowledge, and concussion education experiences. The aim of this investigation is to explore the following in-depth research questions regarding the results of the survey:

1. What is the difference between male and female respondents on knowledge of concussion definition and symptom identification?
2. Do collegiate males and females differ on knowledge of available concussion support services?
3. Do collegiate males and females differ on general concussion knowledge questions?
4. What method of concussion education is most preferred by male vs female respondents?
5. Did the number of the concussions reported change from the beginning to the end of the survey?

The first four of these research questions will be analyzed on the overall scale and broken down between groups (varsity athletes, recreational athletes, and non-athletes). The third research questions stems from respondents' answers to nine general concussion knowledge questions. These questions are displayed below in table 1, which includes the questions key, question being referenced, and the correct answer for that question. The key is used later in the report for abbreviated identification purposes.

The fifth research question relates to two almost identical survey questions at the beginning and end of the survey. The goal of the fifth research question is to determine if there are significant differences in the way that these questions were answered. More specifically, this research question looks to see if respondents reported different numbers of concussions or mild head injuries as a result of being exposed to the information in the survey.

Key	Question	Answer
Q46	Females often experience a longer recovery period following a mild head injury or concussion than males.	True
Q47	The most commonly experience symptom following a concussion or mild head injury is _____ .	Headache
Q48	In order to be diagnosed with a concussion or mild head injury a person must have lost consciousness (e.g., blacked out).	False
Q49	Up to _____ % of individuals experience long-term deficits following a concussion.	30
Q50	What is the most effective method for diagnosing a concussion or mild head injury?	Neurocognitive testing
Q51	Approximately _____ individuals experience a mild head injury or concussion each year in the United States.	2,000,000
Q52	Experiencing repeated concussions in a short time frame (i.e, hours, days, or weeks) can result in death or prolonged disability/symptoms.	True
Q53	People who have multiple concussions or mild head injuries across their life are more likely to have long-term cognitive, physical, or socio-emotional damages than people who only have one concussion or mild head injury.	True
Q54	On average, it takes approximately 2-3 days to recover following a concussion or mild head injury.	False

Table 1. Key to General Concussion Knowledge Questions.

3. Methods

3.1 Data Cleaning & Importation

To evaluate the concussion knowledge differences among college males and females and their preferences for subsequent concussion education methods, this analysis took a number of holistic perspectives and direct statistical methods to draw its conclusions. The dataset was first cleaned by omitting one respondent who did not provide gender information.

Once sufficiently cleaned to the best of our ability, the data was imported into R Statistical Software (Version 3.4+) using the ‘readxl’ package. Subsequent vizuationizations and analyses were performed in R.

3.2 Data Visualizations

Data visualizations we constructed to compare patterns and distributions on many of the research questions of interest. These visualizations aided in our holistic approach to the research.

3.3 Chi-square Test for Independence

To investigate the differences of general concussion knowledge between males and females, a Chi-square test for independence is conducted on nine questions of the survey that pertain to general concussion knowledge. The questions referred to in this section are Q46-Q54 of the survey. Some of the questions follow a true/false format, while others are multiple choice. Each of these concussion knowledge questions were treated as correct/incorrect, as determined by a coded answer key that was provided by the researchers. Refer to table 1 for a key of the questions in reference.The hypotheses of the Chi-Square test for Independence are as follows:

H_o: Gender and correctness are independent.

H_a: Gender and correctness are not independent.

The Chi-Square test for independence is appropriate when the sampling method is random, the variables are categorical, and the frequency for each cell in the contingency table is at least five. We assume the data were collected randomly, and the categorical variables of gender (male/female) and correctness (correct, incorrect, and unsure) assuage the second condition. The sample size is large enough that we may proceed with relative confidence.

3.4 Two-way ANOVA

To evaluate the impact of gender and athletic participation on general concussion knowledge, a Two-Way Analysis of Variance (ANOVA) is performed on the number of correct questions out of 9 answered by respondents. ANOVA will illuminate the marginal and joint effects of gender and athletic participation on the total general concussion knowledge of each respondent. Conclusions based on p-value will be taken holistically. The data is taken from the Qualtrics survey is an observational study, and has nonorthogonal data, thus a Type III ANOVA will be analyzed. The general form of the ANOVA model is:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

Where α_i represents an indicator variable for gender, β_j represents an indicator variable for athletic participation, and $(\alpha\beta)_{ij}$ represents the interaction for the two factors. The errors are assumed to be independent, identical, and normally distributed with mean zero.

The response in this test is the total number of questions answered correctly by the respondents, which is calculated by creating a new variable in the dataset that accumulates the number of questions answered correctly in the general concussion knowledge section of the survey. The categorical factors analyzed are gender (male/female) and athletic participation (varsity, club/rec, & none). In order for our ANOVA to be salient, assumptions of normality, independence, and constant variance of errors, as well as the appropriate removal of outliers in the data, must be satisfied. Diagnostic output is given in the appendix, but all assumptions appear to be met, and the ANOVA will be taken with confidence.

3.5 Paired t-test

A paired t-test statistical procedure is used to answer the research question: “Did the number of concussions reported change from the beginning to the end of the survey?”.

Two nearly identical questions (Q41 and Q29) were asked in two separate sections of the survey. Due to a particularity in the Qualtrics questionnaire system, Q41 chronologically preceded Q29. Q41 is formed by 6 sub-questions. Sub-questions 1 through 6 prompt the respondent to indicate the number of concussions they’ve encountered due to various causes: “sporting injuries” (1), “motor vehicle accidents” (2), “falls” (3), “assaults” (4), “abuse” (5), and “other” (6) (Q41 only). Q29 is formed by 5 sub-questions that correspond the first 5 sub-questions in Q41, while omitting the “other” option. The survey provided space for the total number of concussions reported by respondents, but this statistic was not recorded by the Qualtrics system in the provided dataset. Thus, total number of concussions were calculated as the sum of concussions reported by the various causes provided in the question. The hypotheses tested in the paired t-test are as follows:

H_o: The difference between the number of concussions before and after is zero.

H_o: The difference between the number of concussions before and after is not zero.

The assumptions for the paired t-test requires that the dependent variable is continuous, the data are independent and normally distributed, and that all outliers are appropriately handled. The dependent variable of average concussions is continuous, and we assume the data to be independent and normally distributed due to the Central Limit Theorem.

4. Results

4.1 Concussion Definition and Symptom

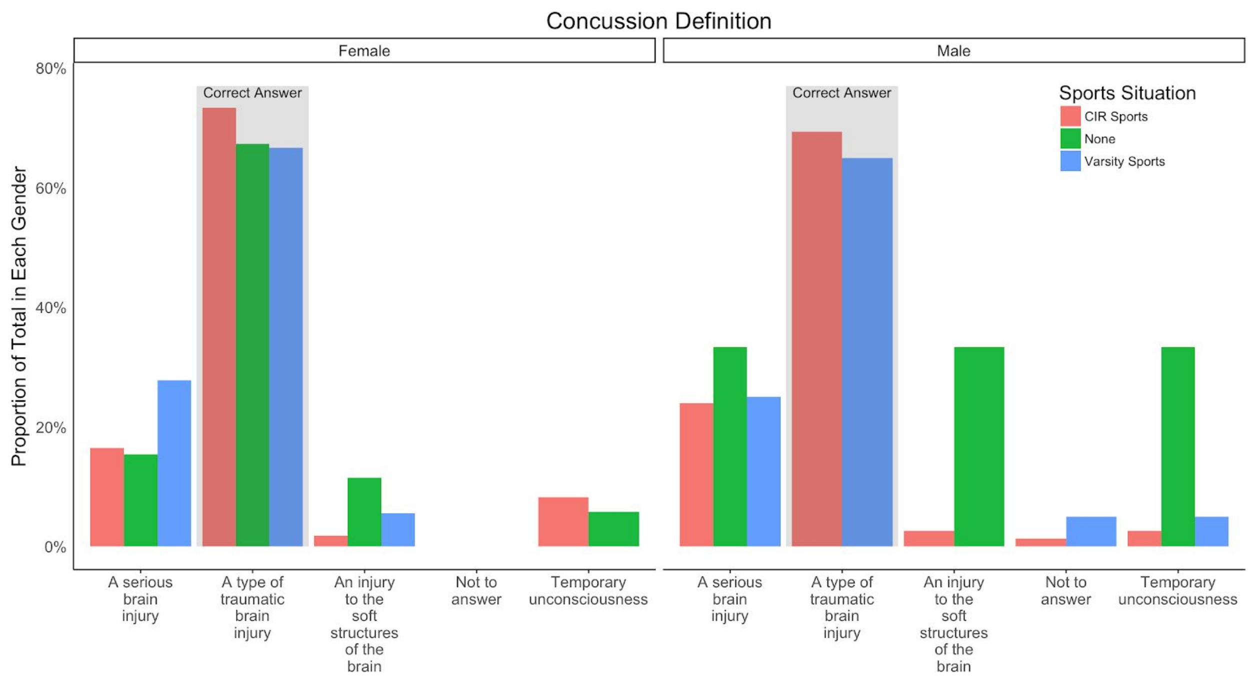

Figure 1. Concussion Definition Distribution by Groups.

As seen above in figure 1, most females and males selected the correct answer for this question on the definition of concussion. A larger proportion of respondents in the recreational athletes category selected the correct answer than those of who are in the group of varsity athletes or non-athletes.

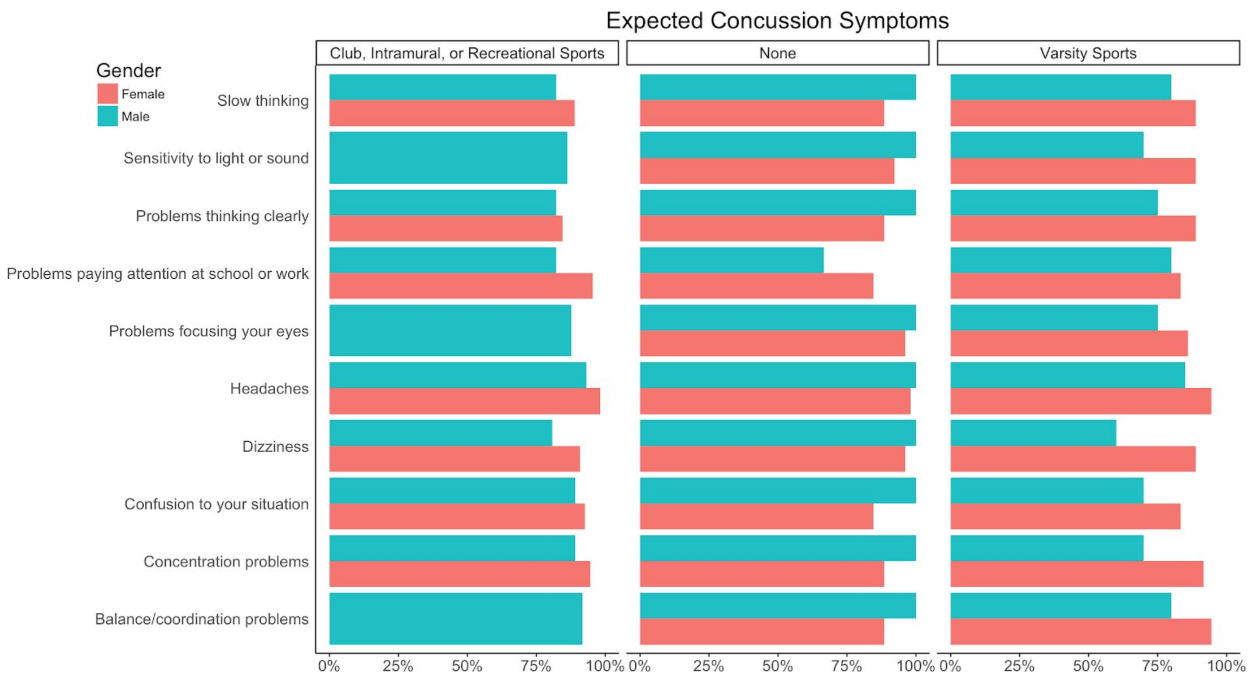

Figure 2. Top 10 Expected Concussion Symptoms by Groups.

Figure 2 contains the top 10 most selected answers for Q76: "*Select all of the symptoms you would expect a person with a concussion or mild head injury to experience*." Respondents hold very similar ideas towards the concussion symptoms. Blindness, deafness or unusual hunger, which are wrong answers, are not included in the top 10 selected answers to the question.

4.2 Concussion Support Services

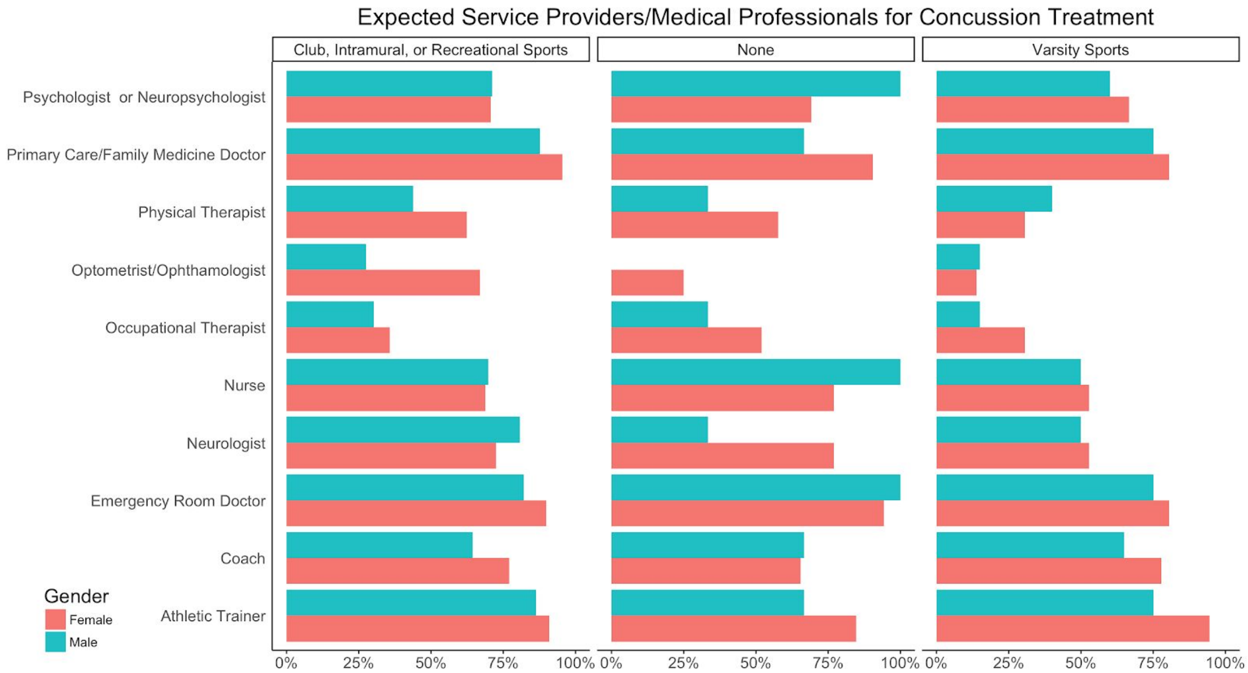

Figure 3. Top 10 Expected Providers/Medical Professionals for Concussion Treatment.

In figure 3, the top 10 most popular service providers/medical professionals for concussion treatment are shown. These selections came from Q73: "*Select all of the service providers/medical professionals that you would expect to be involved in the treatment and recovery of a person with a concussion or mild head injury*." Survey respondents collectively show a higher preference towards athletic trainers, primary care/family medicine doctors, and emergency room doctors. There are, however, some differences in how each athletic/non-athletic group responded to this question broken down by gender. For example, a much higher amount of male non-athletes believe psychologists or neuro-psychologists are expected to be involved in concussion treatment than do female non-athletes.

4.3 General Concussion Knowledge

For the following questions, assumptions that females and males have similar backgrounds of general concussion knowledge are made. The proportions of females and males that selected the correct answer are similar.

Question	Sex	Correct	Incorrect	Not Sure	χ^2
Q46: Do females have a longer recovery time post concussion?	F	48 (24%)	29 (15%)	121 (61%)	1.953
	M	19 (19%)	11 (11%)	68 (70%)	
Q47: Most common post-concussion symptom?	F	145 (73%)	48 (25%)	5 (2%)	0.087
	M	72 (73%)	23 (24%)	3 (3%)	
Q48: Is loss of consciousness needed for diagnosis?	F	189 (98%)	2 (1%)	1 (1%)	0.982
	M	96 (98%)	2 (2%)	0 (0%)	
Q49: % of individuals with long-term deficits following concussion?	F	55 (28%)	83 (42%)	60 (30%)	4.023
	M	21 (21%)	36 (37%)	41 (42%)	
Q50: Best method for concussion diagnosis?	F	90 (45%)	75 (38%)	33 (17%)	0.860
	M	47 (48%)	32 (33%)	19 (19%)	
Q51:How many people have a concussion each year in the US?	F	18 (9%)	101 (51%)	79 (40%)	4.485
	M	17 (17%)	43 (44%)	38 (39%)	
Q52: Can many concussions over a short time cause death?	F	183 (92%)	1 (1%)	14 (7%)	0.346
	M	91 (93%)	1 (1%)	6 (6%)	
Q53: Do multiple concussions lead to long-term effects?	F	177 (89%)	11 (6%)	10 (5%)	3.471
	M	92 (94%)	1 (1%)	5 (5%)	
Q54: Does it take 2-3 days to recover from a concussion?	F	120 (60%)	39 (20%)	39 (20%)	6.551*
	M	74 (76%)	13 (13%)	11 (11%)	

Table 2: Breakdown of survey responses by question.

Ns p>.05
* p≤..05
** p≤.01

Based on the χ^2 value of table 2, females and males only have different opinions for Q54. These two groups hold similar understandings for the rest of questions. Overall, females and males have a very similar general concussion knowledge background.

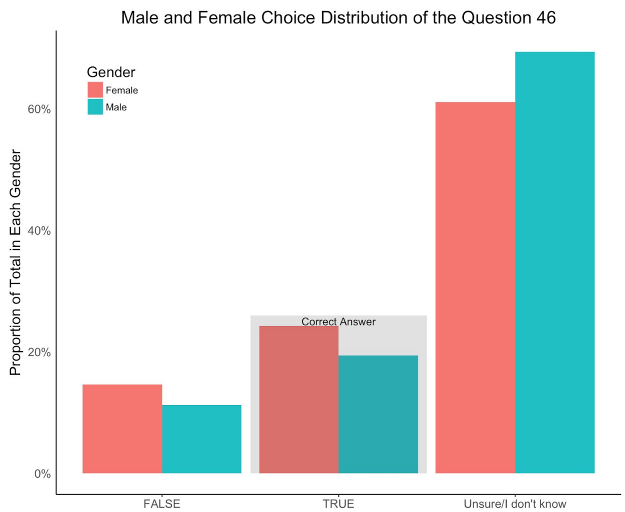

Figure 4. Q46: Do females have a longer recovery time post concussion?

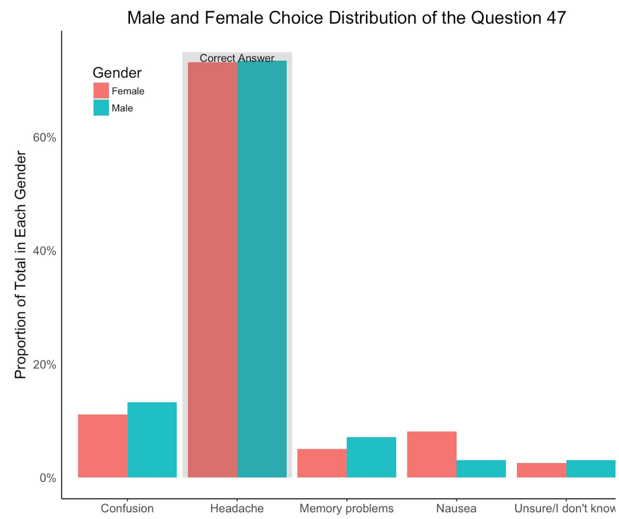

Figure 5. Q47: Most common post-concussion symptom?

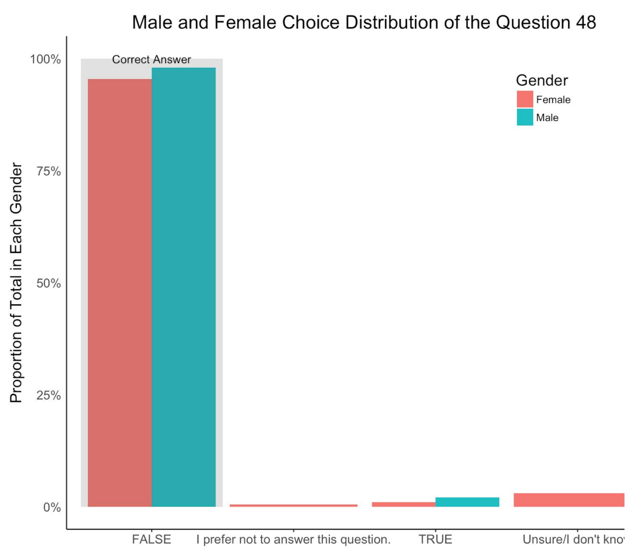

Figure 6. Q48: Is loss of consciousness needed for diagnosis?

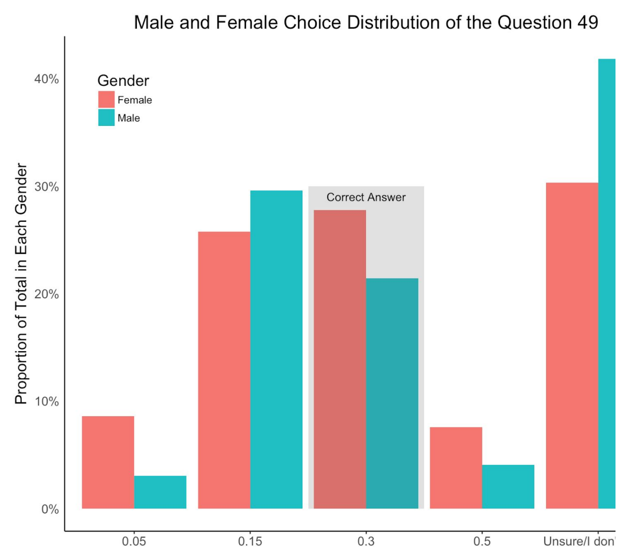

Figure 7. Q49: Percentage of individuals with long-term deficits following concussion?

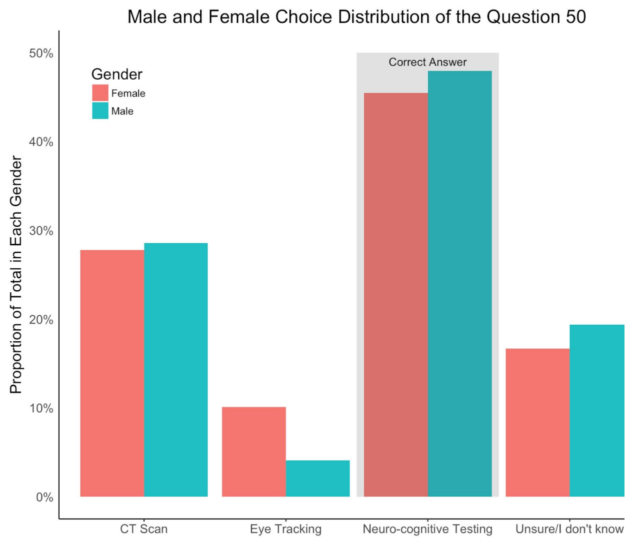

Figure 8. Q50: Best method for concussion diagnosis?

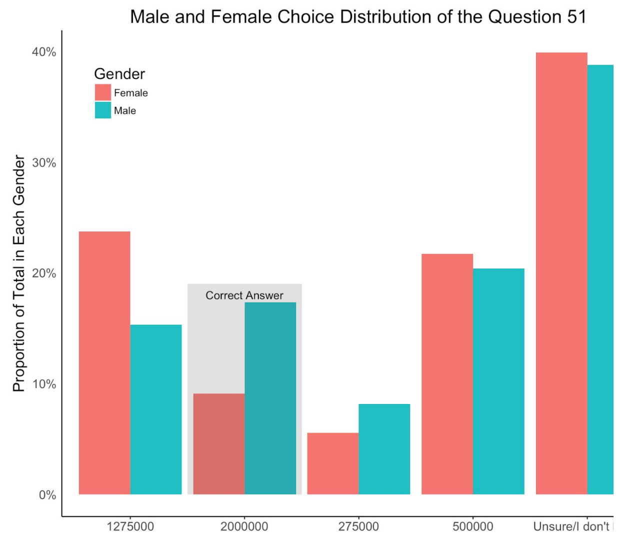

Figure 9. Q51: How many people have a concussion each year in the US?

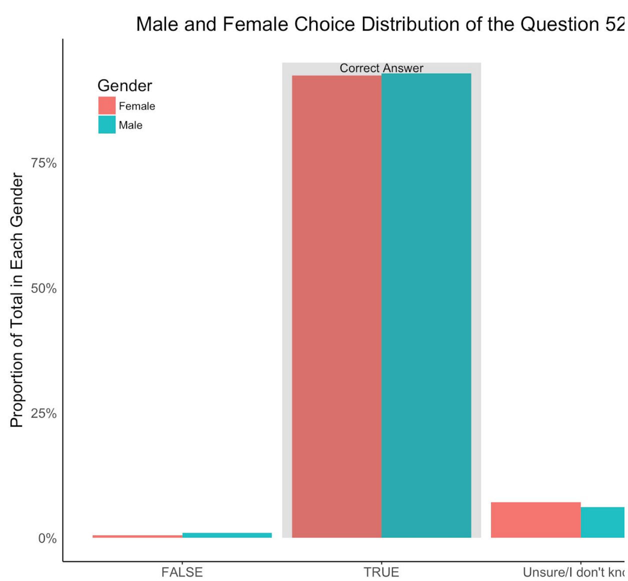

Figure 10. Q52: Can many concussions over a short time cause death?

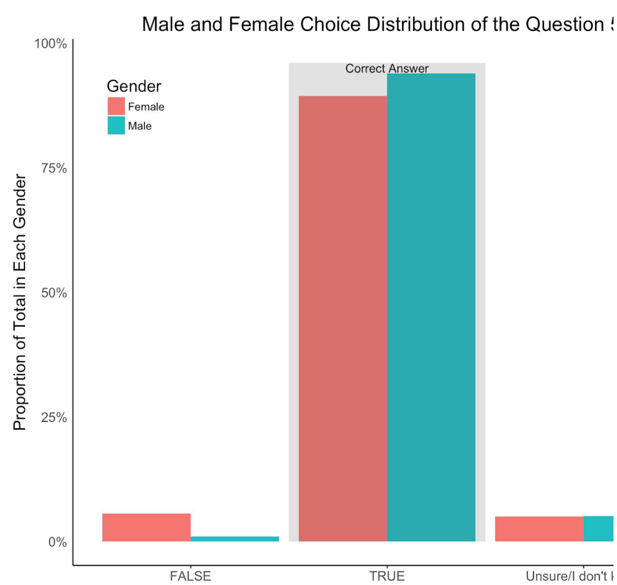

Figure 11. Q53: Do multiple concussions lead to long-term effects?

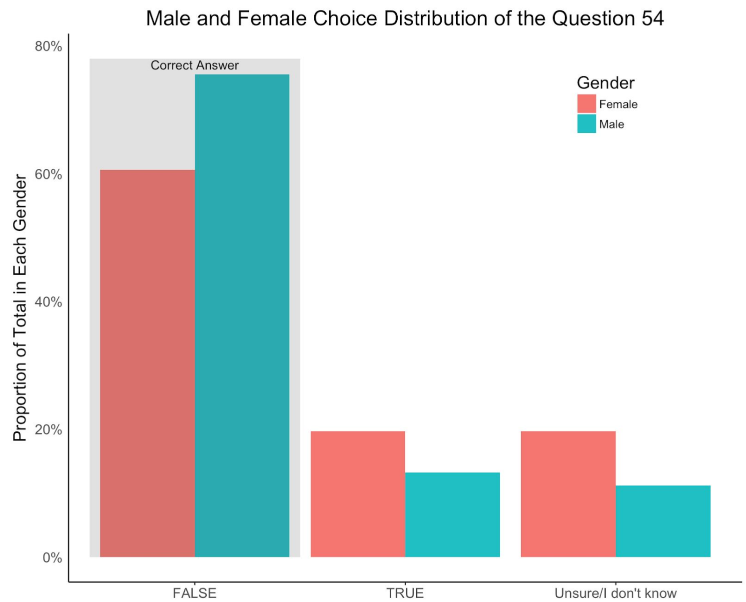

Figure 12. Q54: Does it take 2-3 days to recover from a concussion?

As figures 4 through 12 show, females and males have similar backgrounds of general concussion knowledge, with the exception of Q54 (figure 12), which asks about the recovery time for a concussion or mild head injury. For Q46 (figure 4), which asks about female recovery time compared to male recovery time, a larger proportion of respondents were unsure of the correct answer. For Q49 (figure 7) about the percentage of individuals with long-term deficits following concussion, a large proportion of respondents answered the question incorrectly or were unsure of the answer. For Q50 (figure 8) on the best method for concussion diagnosis, nearly 50% of respondents from both genders chose the correct answer. Still, many incorrectly chose CT scan as the answer. For Q51 (figure 9), asking how many people have a concussion each year in the US, only a small proportion of respondents chose the correct answer and the largest proportion are those that were unsure of the answer.

Group	N	Mean # Correct	Std
Male Varsity Athlete	21	5.50	1.47
Female Varsity Athlete	39	5.67	1.17
Male Club, Rec, Intramural Athlete	77	5.37	1.60
Female Club, Rec, Intramural Athlete	114	5.09	1.41
Male Non-Athlete	3	5.33	.58
Female Non-Athlete	54	5.019	1.38

Table 3. Descriptive statistics for number of concussion knowledge questions correct by gender and athletic participation.

	df	Sum Sq	Mean Sq	F-value	P-value
Gender	1	2.7	2.678	1.764	0.1852
Sports	2	8.5	4.258	2.805	0.0622
Gender:Sports	2	2.9	1.437	0.947	0.3892
Residuals	290	440.3	1.518		

Table 4. Two-way ANOVA results.

Based on the two way ANOVA test result in table 4, there are no significant differences between both genders and sports groups. The p-value for sports is 0.0622 which is higher than the α level (0.05) indicate that is marginal evidence to suggest that there are slight differences between the responses of people who from different sports groups.

4.4 Preferred Concussion Education

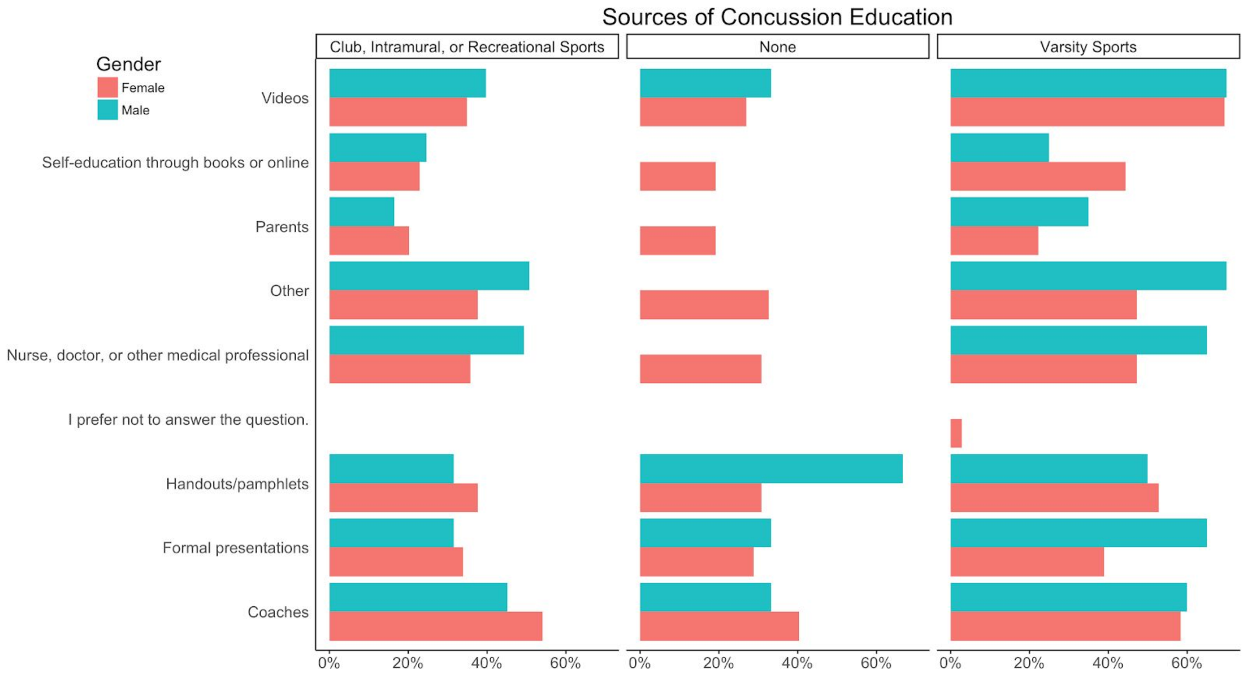

Figure 13. Sources of Concussion Education.

1 = most preferred *7 = least preferred*	**Female-Non-Athletes**	**Female-Rec. Athletes**	**Female-Varsity Athletes**	**Male-Non-Athletes**	**Male-Rec. Athletes**	**Male-Varsity Athletes**
Handouts/Pamphlets	**2.00**	4.24	4.41	4.06	**3.92**	**3.65**
Videos	**3.00**	**2.93**	**2.94**	**3.29**	**3.45**	**2.50**
Formal Presentations	**1.50**	4.03	**3.47**	4.31	4.13	**3.65**
Self-education	4.50	5.00	4.53	5.06	4.88	4.97
Parents	6.00	5.07	5.06	4.63	5.02	5.76
Coaches	6.00	4.11	4.24	**3.88**	**3.76**	4.35
Medical professional	5.00	**2.62**	**3.35**	**2.77**	**2.84**	**3.12**

Table 5. Most prefered concussion education method.

Table 5 refers to Q79: *Drag and drop the numbers to rank your preference for educational method when learning about concussion or mild head injury*. In this question and outlined in Table 5, “1” stands for most preferred and “7” stands for least preferred. The median integer between one and seven is four, so mean preferences that are under 4 indicate a strongly preferred educational method. Table 5 indicates that respondents showed a high preference for learning about concussions through videos no matter which gender or sports group they belong to. The second most popular way to learn about concussions is from medical professionals. Except for female non-athletes, all groups indicated a high preference for learning from medical professionals. Other groups of respondents showed a high preference for learning about concussions by reading informational handouts/pamphlets, watching formal presentations, or through coaches.

4.5 Before and After Comparison

Cause of Concussion	t-Test	DF	P-Value	Confidence Interval
Sporting injuries	-1.31	278	0.1914	(-12.21, 2.45)
Motor Vehicle accidents	0.00	279	1.0000	(-0.03, 0.03)
Falls	-0.58	279	0.5646	(-0.13, 0.07)
Assaults	1.00	279	0.3182	(0.00, 0.01)
Abuse	1.42	279	0.1577	(0.00, 0.02)
Total	-1.31	278	0.1898	(-12.23, 2.44)

Table 6. Paired t-Test Summary For Number of Concussions Reported at Beginning (Q29) and End (Q41) of the Survey.

The paired t-Test results in table 6 indicate that the number of concussions reported did not change from the beginning to the end of the survey. More specifically, the paired t-test shows that the p-value of each sub-question in Q29 corresponding to its’ respective sub-question in Q41 is greater than 0.05, indicating that there is little evidence to suggest a change in the number of concussions reported from the beginning to the end of the survey to the end.

5. General Discussion

The findings of this investigation do not show any meaningful differences between genders in general concussion knowledge. Males and females demonstrated similar levels of correctness from question to question. However, it was seen that some questions were easier to answer than others for this collegiate population. Male and female respondents scored exceedingly well on a few of these general knowledge questions, demonstrating that there is broad awareness of the nature and seriousness of concussions among college students. However, the questions that respondents were mostly unsuccessful in answering, including a question involving differences in concussion recovery patterns between males and females, showed that there is less awareness when it comes to specifics. Overall, our results do not indicate that males and females differ from one another in regards to general concussion knowledge, as they overperform and underperform in tandem. However, some indications are given that more deep learning is required for this college-aged population to better understand the facts surroundings concussions and head injuries.

The survey population was comprised of entirely college students, with a mean age of 20.08 years. This being said, the findings here speak only to the gender differences of college aged students. They do not qualify any other claims regarding gender differences in uneducated populations, or age groups outside of this about five year range.

The findings of the paired t-test indicate that the respondents were consistent in their reported number of concussions/head injuries when prompted in the beginning of the survey, and when prompted again near its completion. This indicates that the survey itself didn’t generally shine any new light on the respondents relationship to concussion knowledge from beginning to end of the survey, at least in a way that caused the respondents to reconsider their previous head injury medical history.

Recommendations for further research on concussion knowledge include broadening the demographic scope of the investigation to span wider socioeconomic, educational, and ethic levels. It is acceptable that concussion knowledge is of utmost importance in the field of sports medicine, and ought to be explored thoroughly and educated aggressively among these demographic ranges. However, if medical professionals and researchers take a more holistic approach to investigating concussion knowledge, they can discover how far concussion knowledge reaches in the general public and hopefully bring awareness to more than just athlete and student populations.

6. References

Wickham, H., & Bryan, J. (2017). readxl: Read Excel Files. R package version 1.0.0. https://CRAN.R-project.org/package=readxl

Wickham, H., Francois, R., Henry, L., & Müller, K. (2017). dplyr: A Grammar of Data Manipulation. R package version 0.7.4. https://CRAN.R-project.org/package=dplyr

Wickham, H. (2009) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2009.

Weisstein, E. W. “ANOVA.” From *MathWorld* -- A Wolfram Web Resource. http://mathworld.wolfram.com/ANOVA/html

Weisstein, E. W. “Chi-Squared Test.” From *Mathworld* -- A Wolfram Web Resource. http://mathworld.wolfram.com/Chi-SquaredTest.html

Weisstein, E. W. “Paired t-Test.” From *MathWorld* -- A Wolfram Web Resource. http://mathworld.wolfram.com/Pairedt-Test.html

7. Appendix

7.1 Assumptions Check

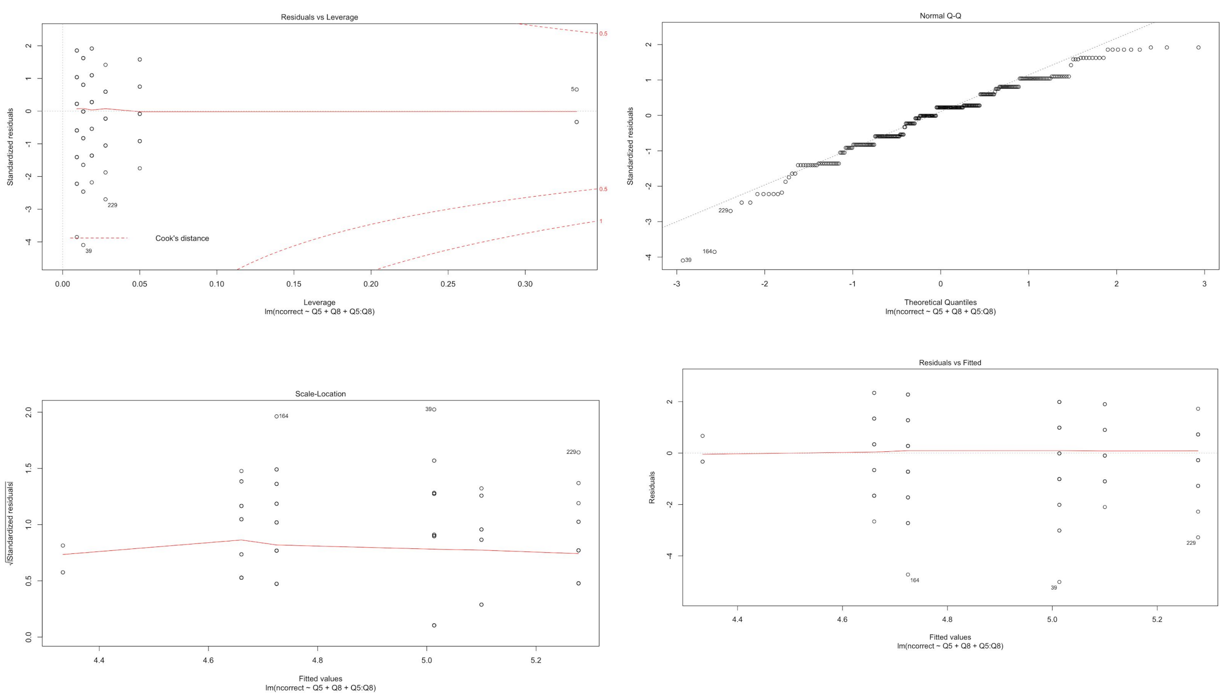

Figure 14. Diagnostic output for the Two-Way ANOVA.

Based on the four figures above, the two way ANOVA model meet the assumptions of constant variance, independence, normality and linearity.

7.2 R Code

```
library(readxl)
library(dplyr)
library(ggplot2)
wave2<-read_xlsx("/Users/Qiu/Documents/STA 475/Project
3/male-female.xlsx")
View(wave2)
##### Some Descriptive Stats ######
wave2$ncorrect <- (wave2$Q46 == "TRUE") + (wave2$Q47 == "Headache") +
(wave2$Q48 == "FALSE") + (wave2$Q49 == "0.3") + (wave2$Q50 ==
"Neuro-cognitive Testing") + (wave2$Q51 == "2000000") + (wave2$Q52 ==
"TRUE") + (wave2$Q53 == "TRUE") + (wave2$Q54 == "FALSE")
```

```
ncorrect <- (wave2$Q46 == "TRUE") + (wave2$Q47 == "Headache") +
(wave2$Q48 == "FALSE") + (wave2$Q49 == "30%") + (wave2$Q50 ==
"Neuro-cognitive testing") + (wave2$Q51 == "2,000,000") + (wave2$Q52
== "TRUE") + (wave2$Q53 == "TRUE") + (wave2$Q54 == "FALSE")
ncorrect
summary(ncorrect)

varsity_males <- wave2[which(wave2$Q5=="Male" & wave2$Q8 =="Varsity
Sports"),]
varsity_females <- wave2[which(wave2$Q5=="Female" &
wave2$Q8=="Varsity Sports"),]
rec_males <- wave2[which(wave2$Q5=="Male" & wave2$Q8 =="Club,
Intramural, or Recreational Sports"),]
rec_females <- wave2[which(wave2$Q5=="Female" & wave2$Q8 =="Club,
Intramural, or Recreational Sports"),]
nonathl_male <- wave2[which(wave2$Q5=="Male" & wave2$Q8 =="None"),]
nonathl_female <- wave2[which(wave2$Q5=="Female" & wave2$Q8
=="None"),]

summary(varsity_males$ncorrect)
summary(varsity_females$ncorrect)
summary(rec_males$ncorrect)
summary(rec_females$ncorrect)
summary(nonathl_male$ncorrect)
summary(nonathl_female$ncorrect)

sd(varsity_males$ncorrect, na.rm = TRUE)
sd(varsity_females$ncorrect, na.rm = TRUE)
sd(rec_males$ncorrect, na.rm = TRUE)
sd(rec_females$ncorrect, na.rm = TRUE)
sd(nonathl_male$ncorrect, na.rm = TRUE)
sd(nonathl_female$ncorrect, na.rm = TRUE)
#46-true
summary(wave2$Q46)
fortysix<-wave2 %>%
  select(StartDate,Q5,Q46)%>%
  group_by(Q5,Q46)%>%
  mutate(PercentageofGroups=sum(!is.na(Q46)))%>%
  summarise(PercentageChoose=length(Q46))
Q46<-as.numeric(fortysix$Q46)
fortysix<-na.omit(fortysix)
View(fortysix)
fortysix<-fortysix%>%
```

```
  group_by(Q5)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))
#datavisual
ggplot(fortysix, aes(x=Q46, y=PercentageChoice, fill=Q5)) +
  geom_col(position="dodge") +
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(0.1,0.87),
        legend.text=element_text(size=10),
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Whether Female have longer recovery period")+
  annotate("text", x = 2, y = 0.25, label = "Correct Answer")+
  annotate("rect", xmin = 1.5, xmax = 2.5, ymin = 0, ymax = 0.26,
           alpha = .2)
#enter the data
male<-c(19,11,68)
female<-c(48,29,121)
fortysix1<-as.data.frame(rbind(male,female))
View(fortysix1)
names(fortysix1)<-c('Right','Wrong','Not sure')
chisq.test(fortysix1)
################################
#47-2-headache
summary(wave2$Q47)
fortyseven<-wave2 %>%
  select(StartDate,Q5,Q47)%>%
  group_by(Q5,Q47)%>%
  mutate(PercentageofGroups=sum(!is.na(Q47)))%>%
  summarise(PercentageChoose=length(Q47))
Q47<-as.numeric(fortyseven$Q47)
fortyseven<-na.omit(fortyseven)
View(fortyseven)
fortyseven<-fortyseven%>%
  group_by(Q5)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))
#datavisual
ggplot(fortyseven, aes(x=Q47, y=PercentageChoice, fill=Q5)) +
  geom_col(position="dodge") +
```

```
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(0.1,0.87),
        legend.text=element_text(size=10),
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Most Common")+
  annotate("text", x = 2, y = 0.74, label = "Correct Answer")+
  annotate("rect", xmin = 1.5, xmax = 2.5, ymin = 0, ymax = 0.75,
           alpha = .2)
female1<-c(145,48,5)
male1<-c(72,23,3)
fortyseven1<-as.data.frame(rbind(male1,female1))
View(fortyseven1)
names(fortyseven1)<-c('Right','Wrong','Not sure')
chisq.test(fortyseven1)

########################################

#48-6-FALSE
summary(wave2$Q48)
fortyeight<-wave2 %>%
  select(StartDate,Q5,Q48)%>%
  group_by(Q5,Q48)%>%
  mutate(PercentageofGroups=sum(!is.na(Q48)))%>%
  summarise(PercentageChoose=length(Q48))
Q48<-as.numeric(fortyeight$Q48)
fortyeight<-na.omit(fortyeight)
View(fortyeight)
fortyeight<-fortyeight%>%
  group_by(Q5)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))
#datavisual
ggplot(fortyeight, aes(x=Q48, y=PercentageChoice, fill=Q5)) +
  geom_col(position="dodge") +
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(0.8,0.87),
```

```
        legend.text=element_text(size=10),
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Male and Female Choice Distribution of the Question 48")+
  annotate("text", x = 1, y = 0.999, label = "Correct Answer")+
  annotate("rect", xmin = 0.5, xmax = 1.5, ymin = 0, ymax = 1,
           alpha = .2)

#enter the data
female2<-c(189,2,1)
male2<-c(96,2,0)
fortyeight1<-as.data.frame(rbind(male2,female2))
View(fortyeight1)
names(fortyeight1)<-c('Right','Wrong','Not Sure')
chisq.test(fortyeight1)

###################################################

#49-3-30%

summary(wave2$Q49)
fortynine<-wave2 %>%
  select(StartDate,Q5,Q49)%>%
  group_by(Q5,Q49)%>%
  mutate(PercentageofGroups=sum(!is.na(Q49)))%>%
  summarise(PercentageChoose=length(Q49))
Q49<-as.numeric(fortynine$Q49)
fortynine<-na.omit(fortynine)
View(fortynine)
fortynine<-fortynine%>%
  group_by(Q5)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))
#datavisual
ggplot(fortynine, aes(x=Q49, y=PercentageChoice, fill=Q5)) +
  geom_col(position="dodge") +
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(0.1,0.87),
        legend.text=element_text(size=10),
```

```
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Male and Female Choice Distribution of the Question 49")+
  annotate("text", x = 3, y = 0.29, label = "Correct Answer")+
  annotate("rect", xmin = 2.5, xmax = 3.5, ymin = 0, ymax = 0.3,
           alpha = .2)
#enter the data
female3<-c(55,83,60)
male3<-c(21,36,41)
fortynine1<-as.data.frame(rbind(male3,female3))
View(fortynine1)
names(fortynine1)<-c('Right','Wrong','Not sure')
chisq.test(fortynine1)
```

```
#########################-----------------------------------------
#50-1-o   Neuro-cognitive Testing
summary(wave2$Q50)
fifty<-wave2 %>%
  select(StartDate,Q5,Q50)%>%
  group_by(Q5,Q50)%>%
  mutate(PercentageofGroups=sum(!is.na(Q50)))%>%
  summarise(PercentageChoose=length(Q50))
Q50<-as.numeric(fifty$Q50)
fifty<-na.omit(fifty)
fifty<-fifty%>%
  group_by(Q5)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))
View(fifty)
#datavisual
ggplot(fifty, aes(x=Q50, y=PercentageChoice, fill=Q5)) +
  geom_col(position="dodge") +
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(0.1,0.87),
```

```
        legend.text=element_text(size=10),
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Male and Female Choice Distribution of the Question 50")+
  annotate("text", x = 3, y = 0.49, label = "Correct Answer")+
  annotate("rect", xmin = 2.5, xmax = 3.5, ymin = 0, ymax = 0.5,
           alpha = .2)
#enter the data
female4<-c(90,75,33)
male4<-c(47,32,19)
fifty1<-as.data.frame(rbind(male4,female4))
names(fifty1)<-c('Right','Wrong','Not sure')
View(fifty1)
chisq.test(fifty1)
###################_------------------------------------------------
---



#51--o    2,000,000
summary(wave2$Q51)
fiftyone<-wave2 %>%
  select(StartDate,Q5,Q51)%>%
  group_by(Q5,Q51)%>%
  mutate(PercentageofGroups=sum(!is.na(Q51)))%>%
  summarise(PercentageChoose=length(Q51))
Q51<-as.numeric(fiftyone$Q51)
fiftyone<-na.omit(fiftyone)
fiftyone<-fiftyone%>%
  group_by(Q5)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))
View(fiftyone)
#datavisual
ggplot(fiftyone, aes(x=Q51, y=PercentageChoice, fill=Q5)) +
  geom_col(position="dodge") +
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
```

```
        legend.position = c(0.1,0.87),
        legend.text=element_text(size=10),
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Male and Female Choice Distribution of the Question 51")+
  annotate("text", x = 2, y = 0.18, label = "Correct Answer")+
  annotate("rect", xmin = 1.5, xmax = 2.5, ymin = 0, ymax = 0.19,
           alpha = .2)

#enter the data
female5<-c(18,101,79)
male5<-c(17,43,38)
fiftyone1<-as.data.frame(rbind(male5,female5))
names(fiftyone1)<-c('Right','Wrong','Not sure')
View(fiftyone1)
chisq.test(fiftyone1)
```

```
#52-1o    TRUE
summary(wave2$Q52)
fiftytwo<-wave2 %>%
  select(StartDate,Q5,Q52)%>%
  group_by(Q5,Q52)%>%
  mutate(PercentageofGroups=sum(!is.na(Q52)))%>%
  summarise(PercentageChoose=length(Q52))
Q52<-as.numeric(fiftytwo$Q52)
fiftytwo<-na.omit(fiftytwo)
fiftytwo<-fiftytwo%>%
  group_by(Q5)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))
View(fiftytwo)
#datavisual
ggplot(fiftytwo, aes(x=Q52, y=PercentageChoice, fill=Q5)) +
  geom_col(position="dodge") +
  theme_classic()+
```

```
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(0.1,0.87),
        legend.text=element_text(size=10),
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Male and Female Choice Distribution of the Question 52")+
  annotate("text", x = 2, y = 0.94, label = "Correct Answer")+
  annotate("rect", xmin = 1.5, xmax = 2.5, ymin = 0, ymax = 0.95,
           alpha = .2)
#enter the data
female6<-c(183,1,14)
male6<-c(91,1,6)
fiftytwo1<-as.data.frame(rbind(male6,female6))
names(fiftytwo1)<-c('Right','Wrong','Not sure')
View(fiftytwo1)
chisq.test(fiftytwo1)

#53-1o     TRUE
summary(wave2$Q53)
fiftythree<-wave2 %>%
  select(StartDate,Q5,Q53)%>%
  group_by(Q5,Q53)%>%
  mutate(PercentageofGroups=sum(!is.na(Q53)))%>%
  summarise(PercentageChoose=length(Q53))
Q53<-as.numeric(fiftythree$Q53)
fiftythree<-na.omit(fiftythree)
fiftythree<-fiftythree%>%
  group_by(Q5)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))
View(fiftythree)
#datavisual
ggplot(fiftythree, aes(x=Q53, y=PercentageChoice, fill=Q5)) +
  geom_col(position="dodge") +
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(0.1,0.87),
        legend.text=element_text(size=10),
        text=element_text(size=15),
```

```
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Male and Female Choice Distribution of the Question 53")+
  annotate("text", x = 2, y = 0.95, label = "Correct Answer")+
  annotate("rect", xmin = 1.5, xmax = 2.5, ymin = 0, ymax = 0.96,
           alpha = .2)
#enter the data
female7<-c(177,11,10)
male7<-c(92,1,5)
fiftythree1<-as.data.frame(rbind(male7,female7))
names(fiftythree1)<-c('Right','Wrong','Not sure')
View(fiftythree1)
chisq.test(fiftythree1)
```

```
#54-2o    FALSE
summary(wave2$Q54)
fiftyfour<-wave2 %>%
  select(StartDate,Q5,Q54)%>%
  group_by(Q5,Q54)%>%
  mutate(PercentageofGroups=sum(!is.na(Q54)))%>%
  summarise(PercentageChoose=length(Q54))
Q54<-as.numeric(fiftyfour$Q54)
fiftyfour<-na.omit(fiftyfour)
fiftyfour<-fiftyfour%>%
  group_by(Q5)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))
View(fiftyfour)
#datavisual
ggplot(fiftyfour, aes(x=Q54, y=PercentageChoice, fill=Q5)) +
  geom_col(position="dodge") +
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(0.8,0.87),
        legend.text=element_text(size=10),
```

```
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Male and Female Choice Distribution of the Question 54")+
  annotate("text", x = 1, y = 0.77, label = "Correct Answer")+
  annotate("rect", xmin = 0.5, xmax = 1.5, ymin = 0, ymax = 0.78,
           alpha = .2)
#enter the data
female7<-c(120,39,39)
male7<-c(74,13,11)
fiftyfour1<-as.data.frame(rbind(male7,female7))
names(fiftyfour1)<-c('Right','Wrong','Not sure')
View(fiftyfour1)
chisq.test(fiftyfour1)


#2-way anova
summary(wave2)
total<-wave2 %>%
  select(Q5,Q8,Q46,Q47,Q48,Q49,Q50,Q51,Q52,Q53,Q54)%>%
  group_by(Q5,Q8,Q46)%>%
  mutate(q46=sum(!is.na(Q46)))%>%
  mutate(q47=sum(!is.na(Q47)))%>%
  mutate(q48=sum(!is.na(Q48)))%>%
  mutate(q49=sum(!is.na(Q49)))%>%
  mutate(q50=sum(!is.na(Q50)))%>%
  mutate(q51=sum(!is.na(Q51)))%>%
  mutate(q52=sum(!is.na(Q52)))%>%
  mutate(q53=sum(!is.na(Q53)))%>%
  mutate(q54=sum(!is.na(Q54)))


total<-total[!duplicated(total$q46),]
total<-total[!duplicated(total$q47),]
total<-total[!duplicated(total$q48),]
total<-total[!duplicated(total$q49),]
total<-total[!duplicated(total$q50),]
total<-total[!duplicated(total$q51),]
total<-total[!duplicated(total$q52),]
total<-total[!duplicated(total$q53),]
total<-total[!duplicated(total$q54),]
```

```
View(total)
  mutate(PercentageofGroups=sum(!is.na(Q46)))%>%
  summarise(PercentageChoose=length(Q46))
Q46<-as.numeric(fortysix$Q46)
fortysix<-na.omit(fortysix)
View(fortysix)
fortysix<-fortysix%>%
  group_by(Q5)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))




total<-wave2 %>%
  select(Q5,Q8,Q54)%>%
  group_by(Q5,Q8,Q54)%>%
  mutate(qq=sum(!is.na(Q54)))

total<-total[!duplicated(total$qq),]

total1<-total%>%
  group_by(Q5,Q8)%>%
  mutate(PercentageChoice=qq/sum(qq))


View(total1)

#male=1,female=2, varisity=1,club=2,none=3

ano<-aov(ncorrect~Q5+Q8+Q5:Q8,data=wave2)
summary(ano)
lm11<-lm(ncorrect~Q5+Q8+Q5:Q8,data=wave2)
summary(lm11)
plot(lm11)



table(anova$CorrectRate)
View(wave2)
```

```
#definition
definition<-wave2 %>%
  select(Q5,Q8,Q74)%>%
  group_by(Q5,Q8,Q74)%>%
  mutate(PercentageofGroups=sum(!is.na(Q74)))%>%
  summarise(PercentageChoose=length(Q74))
Q74<-as.numeric(definition$Q74)
definition<-na.omit(definition)
View(definition)
definition<-definition%>%
  group_by(Q5,Q8)%>%
  mutate(PercentageChoice=PercentageChoose/sum(PercentageChoose))
View(definition)
definition$Category[definition$Q8=="Club, Intramural, or Recreational
Sports"]<-"CIR Sports"
definition$Category[definition$Q8=="None"]<-"None"
definition$Category[definition$Q8=="Varsity Sports"]<-"Varsity
Sports"
definition$cat[definition$Q74=="Temporary unconsciousness caused by a
blow to the head."]<-"Temporary unconsciousness"
definition$cat[definition$Q74=="An injury to the soft structures of
the brain causing brief memory loss."]<-"An injury to the soft
structures of the brain"
definition$cat[definition$Q74=="A type of traumatic brain injury
caused by a bump, jolt, or blow to the head."]<-"A type of traumatic
brain injury"
definition$cat[definition$Q74=="A serious brain injury caused by back
and forth motion of the brain in the skull."]<-"A serious brain
injury"
definition$cat[definition$Q74=="I prefer not to answer the
question."]<-"Not to answer"
get_wraper <- function(width) {
  function(x) {
    lapply(strwrap(x, width = width, simplify = FALSE), paste,
collapse="\n")
  }
}
ggplot(definition, aes(x=cat, y=PercentageChoice, fill=Category)) +
  geom_col(position="dodge") +
  theme_classic()+
```

```
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(0.9,0.87),
        legend.text=element_text(size=10),
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Sports Situation"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  #  scale_x_discrete(labels=c("1" = "Very Bad", "2" = "2-Bad",
  #                            "3" = "3-Fair","4"="4-Good","5"="5-Very
Good"))+
  ggtitle("Male and Female Differences on Knowledge of Concussion
Definition")+
  facet_grid(~ Q5)+
  scale_x_discrete(labels = get_wraper(10))+
  annotate("text", x = 2, y = 0.76, label = "Correct Answer")+
  annotate("rect", xmin = 1.5, xmax = 2.5, ymin = 0, ymax = 0.77,
           alpha = .2)

#####
Q39<-read_xlsx("/Users/Qiu/Documents/STA 475/Project 3/Q39.xlsx")
View(Q39)
ggplot(Q39, aes(x=Choice, y=Porportion, fill=Gender)) +
  geom_col(position="dodge") +
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        legend.position = c(-0.2,0.95),
        legend.text=element_text(size=10),
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Male and Female Differences on Knowledge of Concussion
Definition")+
  facet_grid(~ Sports)+coord_flip()

###39
Q39<-read_xlsx("/Users/Qiu/Documents/STA 475/Project 3/Q39.xlsx")
View(Q39)
ggplot(Q39, aes(x=Choice, y=Porportion, fill=Gender)) +
```

```
  geom_col(position="dodge") +
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        axis.title.y = element_blank(),
        legend.position = c(-0.2,0.95),
        legend.text=element_text(size=10),
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Sources of Concussion Education")+
  facet_grid(~ Sports)+coord_flip()


###73
Q73<-read_xlsx("/Users/Qiu/Documents/STA 475/Project
3/top10-Q73.xlsx")
View(Q73)
ggplot(Q73, aes(x=Choice, y=Porportion, fill=Gender)) +
  geom_col(position="dodge") +
  theme_classic()+
  theme(axis.title.x = element_blank(),
        axis.ticks.y = element_blank(),
        axis.title.y = element_blank(),
        legend.position = c(-0.2,0.05),
        legend.text=element_text(size=10),
        text=element_text(size=15),
        plot.title = element_text(hjust = 0.5))+
  guides(fill=guide_legend(title="Gender"))+
  scale_y_continuous(labels = percent)+
  labs(y="Proportion of Total in Each Gender")+
  ggtitle("Expected Service Providers/Medical Professionals for
Concussion Treatment")+
  facet_grid(~ Sports)+coord_flip()


###76
Q76<-read_xlsx("/Users/Qiu/Documents/STA 475/Project
3/top10-Q76.xlsx")
View(Q76)
ggplot(Q76, aes(x=Choice, y=Porportion, fill=Gender)) +
```

```
geom_col(position="dodge") +
theme_classic()+
theme(axis.title.x = element_blank(),
      axis.ticks.y = element_blank(),
      axis.title.y = element_blank(),
      legend.position = c(-0.2,0.95),
      legend.text=element_text(size=10),
      text=element_text(size=15),
      plot.title = element_text(hjust = 0.5))+
guides(fill=guide_legend(title="Gender"))+
scale_y_continuous(labels = percent)+
labs(y="Proportion of Total in Each Gender")+
ggtitle("Expected Concussion Symptoms")+
facet_grid(~ Sports)+coord_flip()
```