Supplemental Information for *Prestige or Perish: Publishing Decisions in Academic Archaeology*

Erik Gjesfjeld, Jess Beck, Stephen Chrisomalis

2021-08-18

Introduction

This markdown document presents the R code for the analysis of responses from the Prestige or Perish: Publishing Decisions in Academic Archaeology survey of carried out by Drs. Jess Beck, Erik Gjesfjeld, and Stephen Chrisomalis. The data presented here was derived from a survey that was taken by respondents through the Qualtrics platform with details provided in the manuscript and supplemental information.

Data Sources

The analysis preformed in this research is comprised of six sources of data. These include:

1. Demographic information for each respondent (PDIA_Cleaned.csv)
2. List of Journal names given by each respondent (PDIA_Cleaned.csv)
3. Ranking of prestige factors given by each respondent (PDIA_Cleaned.csv)
4. Comments provided by respondents (PDIA_Comments.csv)
5. Scientific journal rankings (SJRs) as listed by Scimago (Journals_SJR_Ulrichs.csv)
6. Additional Journal metadata as provided Ulrich's database (Journals_SJR_Ulrichs.csv)

Data Cleaning

Data used in this work was initially cleaned by Dr. J. Beck with further cleaningby Dr. E. Gjesfjeld. The cleaning process broadly included the following:
- Removal of timestamp information for respondents
- Standardization of journal names (i.e. removing journal abbreviations) to coincide with Ulrich's journal names
- Summarizing comments

Data Analysis

The organization and analysis of the data is performed using R and these related packages, which can be loaded into your R environment using the following. (Note: For beginning R users, these packages must first be installed using the install.packages() function) As a note, a majority of the code presented uses "tidy" syntax and commands. If users are not familiar with the tidyverse (Wickham et al. 2019), please see the many great online tutorials and cheat sheets.

```r
#install.packages("tidyverse)
#install.packages("ggpubr")
#install.packages("ggrepel")
library(tidyverse)
library(ggpubr)
library(ggrepel)
library(knitr)
```

The two repositories for the data and code can be found at the following links:

Figsahre https://figshare.com/s/24cf4ca08c5f62ea52af and Erik Gjesfjeld's GitHub Repo https://github.com/erikgjes/Publishing_Decisions_in_Archaeology

The previously cleaned survey data can be loaded directly into your R environment with

```r
#Reading in the cleaned data
results <- read.csv('https://git.io/JLA1W')
```

The data collected from Scimago and Ulrich's can be read in with the following command:

```r
#Reading in the journal metadata and SJR scores for journals
journal_scores<-read.csv('https://git.io/JLA14')
```

Demographic Information

We summarized demographic information on respondents using the following code. Here we present the results for gender, age, and career stage.

```r
#Gender results
gender <- results %>%
  drop_na(Gender) %>%
  group_by(Gender) %>%
  summarize(count=n()) %>%
  mutate(prop=round(count/sum(count),digits=2))

#Age results
age <- results %>%
  drop_na(Age) %>%
  group_by(Age) %>%
  summarize(count=n()) %>%
  mutate(prop=round(count/sum(count),digits=2))

#Career stage results
career <- results %>%
  drop_na(Career) %>%
  group_by(Career) %>%
  summarize(count=n()) %>% mutate(prop=round(count/sum(count),digits=2)) %>%
  mutate(Position_Short=c("Asst Prof","Assoc Prof","CRM", "Emeritus",
                          "Professor", "Masters","Other",
                          "PhD","Post-Doc","Visit Prof"))
```

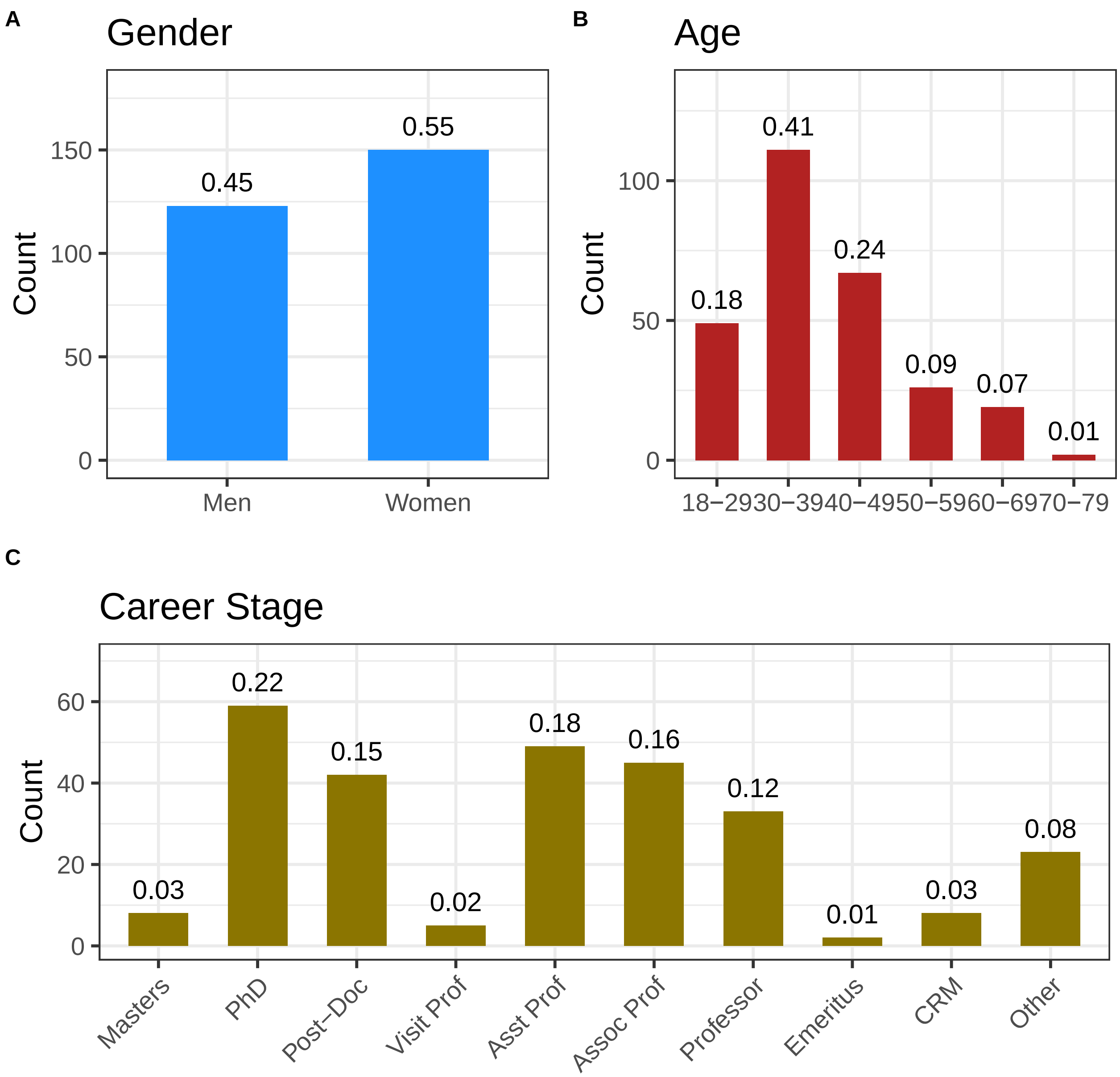
The demographic information can be plotted with the following:

```r
#Gender histogram
gender_hist<-ggplot(gender,aes(x=Gender,y=count)) +
  geom_bar(stat="identity",fill="dodgerblue",width=0.6) +
  labs(x="",y="Count",title="Gender") +
  geom_text(aes(label = prop),
            vjust=-0.75,
            color = "black",
            size = 6) +
    theme_bw(base_size=20) +
  ylim(0,max(gender$count)*1.2) +
  scale_x_discrete(labels = c("Men","Women"))

#Age histogram
age_hist<-ggplot(age,aes(x=Age,y=count)) +
  geom_bar(stat="identity",fill="firebrick",width=0.6) +
  labs(x="",y="Count",title="Age") +
  geom_text(aes(label = prop),
            vjust=-0.75,
            color = "black",
            size = 6) +
    theme_bw(base_size=20) +
  ylim(0,max(age$count)*1.2)

#Reorder the x-axis for career stage following chronological
#order rather than alphabetical order
career_axis <- data.frame(Position_Short=c("PhD", "Asst Prof", "Assoc Prof",
                                           "Post-Doc","Professor",
                                           "Other","CRM","Masters",
                                           "Visit Prof","Emeritus"),
                          car_x_pos=c(2,5,6,3,7,10,9,1,4,8))

#Joining the reordered career stages to the survey data
career_order<- left_join(career, career_axis, by="Position_Short")

#Career histogram
career_hist<-ggplot(career_order,
                    aes(x=reorder(Position_Short,car_x_pos),y=count)) +
  geom_bar(stat="identity",fill="gold4",width=0.6) +
  labs(x="",y="Count",title="Career Stage") +
  geom_text(aes(label = prop),
            vjust=-0.75,
            color = "black",
            size = 6) +
  theme_bw(base_size=20) +
  ylim(0,max(career_order$count)*1.2) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.margin = unit(c(0.5,0.5,0.5,0.5),"cm"))
```

The figure displayed in the manuscript can be created with the following code. Note: This will create a jpeg image file in your working directory.

```r
#Create jpeg file
jpeg("Figure_1.jpg",width=12,height=10,units="in",res=300)
```

```r
#Arrange previous histograms and plot
ggarrange(ggarrange(gender_hist, age_hist, ncol = 2, labels = c("A", "B")),
          career_hist,
          nrow = 2,
          labels = "C",
          vjust = 35.5
          )

dev.off()
```



Results

This research project aimed to examine four key research questions. The data analysis and plotting used to evaluate each question can be found below. Please see the manuscript for additional details.

Question 1

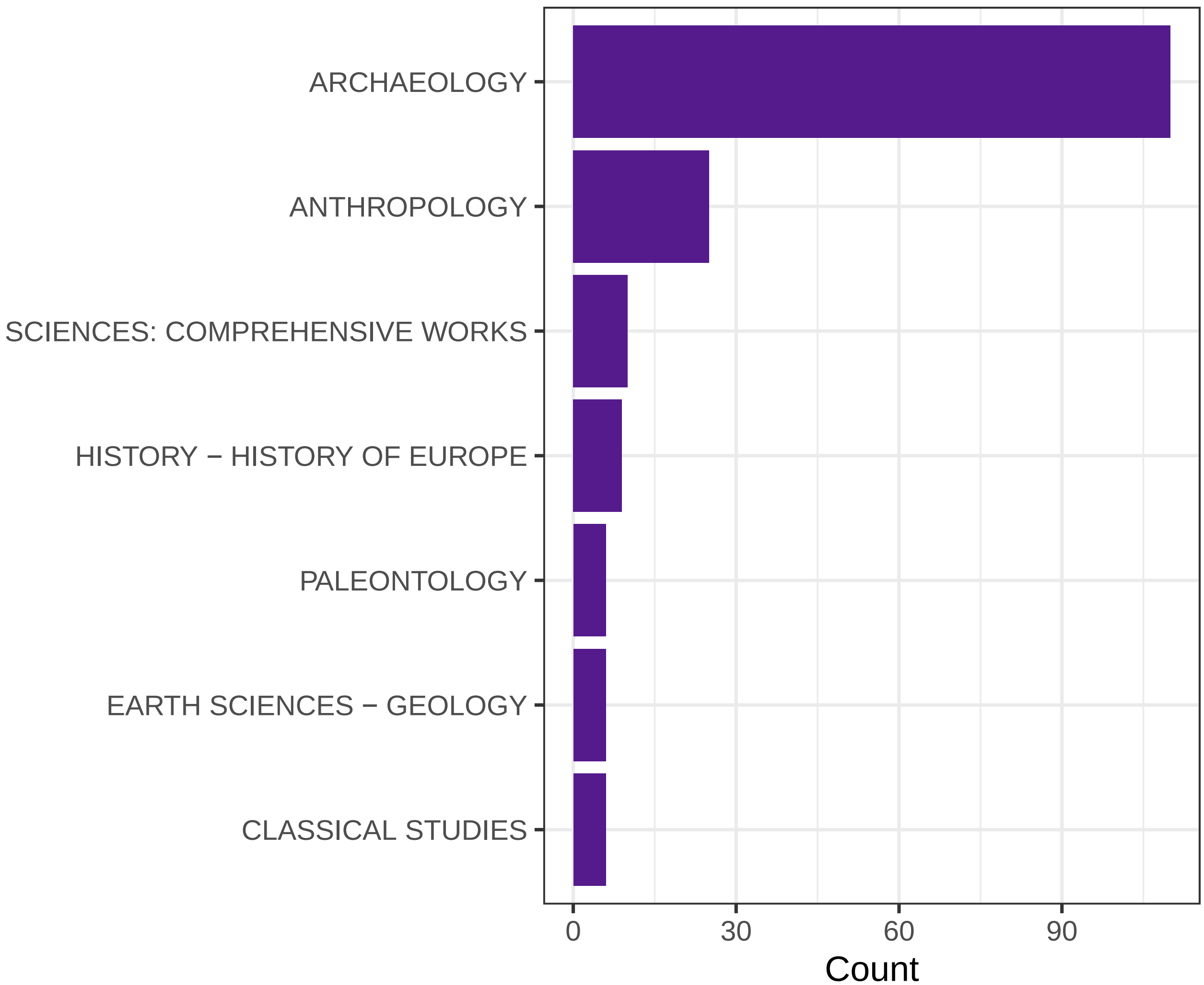
How diverse is the landscape of archaeological publishing?

In order to answer this question, we relied on Ulrich's Serial Directory (UlrichsWeb) classification of subject area for each of the journals listed in the survey results. Each journal can have up to four subject areas with all four subject areas included in our analysis when applicable.

```
#Read in subject areas for listed journals
subject_ulrichs<-read.csv("https://git.io/JLA1u")

#Counting up the subject areas across all four subject areas
subject_table <- as.data.frame(table(c(subject_ulrichs$subject_1,
                                       subject_ulrichs$subject_2,
                                       subject_ulrichs$subject_3,
                                       subject_ulrichs$subject_4))) %>%
  filter(Var1 != "")

#Renaming the columns
colnames(subject_table) <- c("Subject","Count")

#Filter out only subject areas with more than five occurrences
#and arranging them in descending order
subject_table_top <- subject_table %>% arrange(desc(Count)) %>%
  filter(Count > 5)
```

This tally of subject areas can be plotted with the following:

```
#Create jpeg file
jpeg("Figure_2.jpg",width=12,height=7,units="in",res=300)

#First this arranges based on count and then plots as a horizontal bar chart
subject_table_top %>%
  arrange(Count) %>%
  mutate(name=factor(Subject, levels=Subject)) %>%
  ggplot( aes(x=name, y=Count)) +
  geom_col(size=4,fill="purple4") +
  coord_flip() +
  theme_bw(base_size = 20) +
  xlab("")

dev.off()
```



Question 2

To what degree do archaeologists' perceptions of journal prestige match bibliometric journal rankings?

This question required the integration of our survey data with scores from the Scimago Journal Rankings (SJR). The first step was to parse the survey data in order to tally or count how many times each was listed in our data.

```
#Subsetting results for question 1, tallying the results, and calculate rank and proportion of each jou
Q1<-apply(results[,11:16],2,c)
Q1_table<-as.data.frame(table(Q1)) %>% arrange(desc(Freq))
names(Q1_table) <- c("journal_name","count")
Q1_table$prop<-round(Q1_table$count/sum(Q1_table$count),digits=2)
Q1_table$rank <- as.numeric(as.factor(-Q1_table$count))
```

Next, we will also want to merge (or join) each of the journal names given as a response with the Scimago Journal Ranks (SJR) as gathered independently from Scimago.

```r
#Joining survey data and SJR scores based on journal name
Q1_SJR<-left_join(Q1_table,journal_scores,by="journal_name")
#Arranging in descending order and adding a
#rank category (although not used in our analysis)
Q1_SJR <- Q1_SJR %>% arrange(desc(SJR)) %>% mutate(SJR_rank=rank(-SJR)) %>%
  filter(journal_name !="")
```

Figure 3A presents the relationship between SJR scores and the prestige of journals (based on the counts from question 1) and can be visualized with the following:

```r
#Subsetting the journals that have both lower counts and lower SJR scores
Q1_outs<-filter(Q1_SJR, count >= 50 | SJR >= 7) %>% drop_na()
Q1_inset<-filter(Q1_SJR, count <= 50 | SJR <= 7)
#Plotting the main plot and adding labels to commonly listed journals
Q1_plot <- ggplot(Q1_SJR,aes()) +
  geom_point(aes(x=count,y=SJR),color="navyblue",size=3) +
  lims(x=c(-1,200),y=c(-1,30)) +
  labs(x="Count",y="SJR",title="") +
  annotate("rect",xmin=-1,xmax=50,ymin=-1,ymax=7,fill="grey50",
           alpha=0.2,color="grey50") +
  geom_text_repel(data=Q1_outs,aes(x=count,y=SJR,
                             label=c("Cell", "Nature", "Science",
                                     "Proc. Natl. Acad. Sci. U.S.A.",
                                     "J. Archaeol. Sci.", "Curr. Anthropol.",
                                     "J. Anthropol. Archaeol.",
                                     "J. Archaeol. Method Theory",
                                     "Am. Antiq.", "Antiquity")),
                  hjust=c(0.3,0.3,0.3,0.1,0.3,0.3,0.2,0.-0.3,0.3,0.3),
                  vjust=c(-1,-1,-1,-1,-1,-2,-.5,3,2,-1),
                  min.segment.length = 1) +
  theme_bw(base_size=20)
#Subsetting the journals to be plotted in the inset plot
Q1_inset_counts <- filter(Q1_inset, count <= 50 & count >= 38 & SJR <7)
Q1_inset_SJR <- filter(Q1_inset, SJR >= 2.2 & SJR <= 7 & count <=50)
Q1_inset_outs <- rbind(Q1_inset_counts,Q1_inset_SJR)
#Plotting the inset plot along with labels for journals
Q1_inset_plot<-ggplot(Q1_inset,aes()) +
  geom_point(aes(x=count,y=SJR),color="navyblue",size=3) +
  lims(x=c(0,50),y=c(0,7)) +
  labs(x="Count",y="SJR",title="") +
  geom_text_repel(data=Q1_inset_outs,aes(x=count,y=SJR,
                                   label=c("J. Archaeol. Res.",
                                           "World Archaeol.", "Sci. Adv.",
                                           "Nat. Commun.", "Nat. Eco. Evol.",
                                           "Nat. Hum. Behav.", "Quat. Sci. Rev.",
                                           "Proc. R. Soc. B")),
            hjust=c(.3,.3,.3,.7,.1,.1,.3,.3),
            vjust=c(-1,-1,-1,-1,-1,-1,-1,-1),
            min.segment.length = 1) +
            theme_bw(base_size=20) +
              theme( plot.background = element_rect(fill = "transparent",colour = NA),
                     panel.background = element_rect(fill = alpha('grey50',0.2),colour = NULL))
```

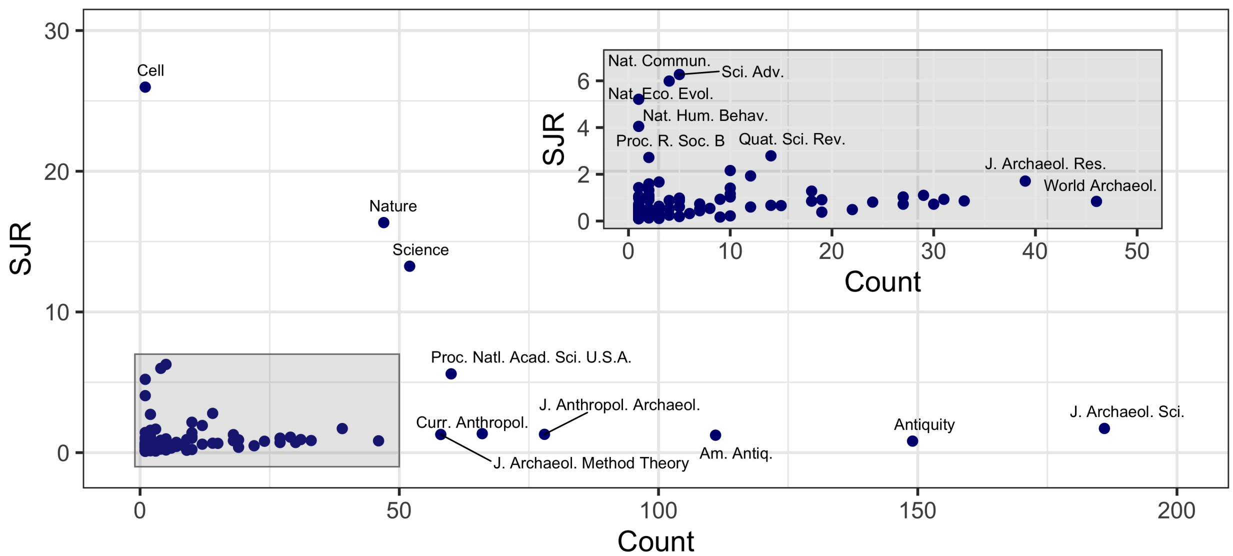
Figure 3B presents the relationship between SJR scores and the prestige of journals (based on the counts from question 1) and can be visualized with the following:

```
#Subsetting the journals that have both lower counts
#and lower SJR scores (Question 2)
Q2<-apply(results[,17:19],2,c)
Q2_table<-as.data.frame(table(Q2)) %>% arrange(desc(Freq))
names(Q2_table) <- c("journal_name","count")
Q2_table$prop<-round(Q2_table$count/sum(Q2_table$count),digits=2)
Q2_table$rank <- as.numeric(as.factor(-Q2_table$count))
#Joining survey data and SJR scores based on journal name
Q2_SJR<-left_join(Q2_table,journal_scores,by="journal_name")
#Arranging in descending order and adding
#a rank category (although not used in our analysis)
Q2_SJR <- Q2_SJR %>% arrange(desc(SJR)) %>% mutate(SJR_rank=rank(-SJR)) %>%
  filter(journal_name !="")
#Subsetting the journals to be plotted in the main plot and the inset
#plot (these are lower values as question 2 only asked for three journals)
Q2_outs<-filter(Q2_SJR, count >= 25 | SJR >= 7) %>% drop_na()
Q2_inset<-filter(Q2_SJR, count <= 25 | SJR <= 7)
Q2_inset_counts <- filter(Q2_inset, count <= 25 & count >= 12 & SJR <7)
Q2_inset_SJR <- filter(Q2_inset, SJR >= 3 & SJR <= 7 & count <=25)
Q2_inset_outs <- rbind(Q2_inset_counts,Q2_inset_SJR)
#Plotting the main  plot along with labels for journals
Q2_plot <- ggplot(Q2_SJR,aes()) +
  geom_point(aes(x=count,y=SJR),color="darkgreen",size=3) +
  lims(x=c(-1,100),y=c(-1,30)) +
  annotate("rect",xmin=-1,xmax=25,ymin=-1,ymax=7,fill="grey50",
           alpha=0.2,color="grey50") +
  geom_text_repel(data=Q2_outs,aes(x=count,y=SJR,label=c("Cell","Nature","Science",
                                                         "Proc. Natl. Acad. Sci. U.S.A.",
                                                         "J. Archaeol. Sci.","Curr. Anthropol.",
                                                         "Am. Antiq.", "Antiquity")),
            hjust=c(0.3,0.3,0.3,0.2,0.3,0.3,0.3,0.3),
            vjust=c(-1,-1,-1,-1,-1,-1,-1,-1),
            min.segment.length = 1) +
  theme_bw(base_size = 20) + labs(x="Count",y="SJR")
#Plotting the inset plot along with labels for journals
Q2_inset_plot<-ggplot(Q2_inset,aes()) +
  geom_point(aes(x=count,y=SJR),color="darkgreen",size=3) +
  geom_text_repel(data=Q2_inset_outs,aes(x=count,y=SJR,
                                         label=c("J. Anthropol. Archaeol.",
                                                 "J. Archaeol. Method Theory",
                                                 "PLOS ONE", "Am. Anthopol.",
                                                 "World Archaeol.","Am. J. Archaeol.",
                                                 "Sci. Adv.","Nat. Commun.")),
            hjust=c(0.65,0.3,0.3,.3,0.3,0.3,0.3,0.3),
            vjust=c(-1,-4,3,-1,0,0,0,0),
            min.segment.length = 1) +
  lims(x=c(0,25),y=c(0,7)) +
  theme_bw(base_size = 20) +
  theme(panel.background = element_rect(fill = alpha('grey50',0.2),
                                        colour = NULL)) +
  labs(x="Count",y="SJR")
```

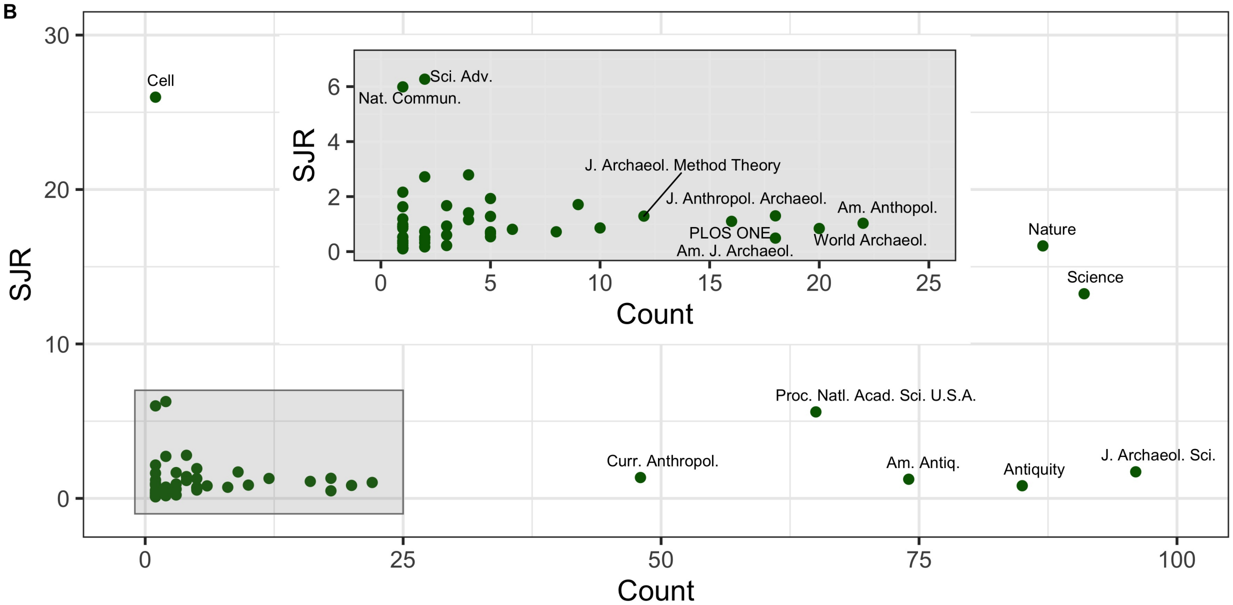
Once these plots have been created the can be arranged and plotted with the following code. It should be noted that the 'ggrepel' package adjusts labels automatically based on screen size. Therefore, it is likely that label placement will be different when rendered on different computers / screens. The original image from the publication is provided for reference which can be found in the GitHub (https://github.com/erikgjes/Prestige-or-Perish-Publishing-Decisions-in-Academic-Archaeology/blob/main/Figure_3.jpg) or FigShare repository.

```
jpeg("Figure_3.jpg",width=12,height=12,units="in",res=300)

par(mfrow=c(2,1))

plot_3A <- Q1_plot + annotation_custom(ggplotGrob(Q1_inset_plot),
                             xmin = 75, xmax = 200, ymin = 10, ymax = 33)

plot_3B <- Q2_plot + annotation_custom(ggplotGrob(Q2_inset_plot),
                             xmin = 13, xmax = 80, ymin = 10, ymax = 30)


ggarrange(plot_3A, plot_3B, ncol=1,nrow=2,
          labels = c("A","B"))


dev.off()
```

A



B



Question 3

How do gender and career stage relate to journals listed to respondents?

In order to examine this question, we aimed to test for significant differences between gender (men and women) and career stage. We further classified career stages into either tenure-track (TT) or non-tenure-track (non-TT). The former includes all respondents who identified as Assistant Professors, Associate Professors, Full Professors, and Emeritus Professors. The latter category (non-TT) includes Masters students, PhD students, Post-Docs, Visiting Professors, CRM/Contract Archaeologists, and any respondents that choose "Other". The division allowed us to perform a Student's t-test for significant differences between the mean SJR scores for our pairs of groups (Men/Women and Tenure-Track/Non-Tenure-Track).

```
#Subsetting response for questions 1 and 2 (again)
q1_responses <- results[,11:16]
q2_responses <- results[,17:19]

#Subsetting the SJR scores and journal names
SJR <- journal_scores$SJR
Names <- journal_scores$journal_name

#Function to match journal names from survey data and from journal score data
f_match<-function(x,new,old){
  new[match(x,old)]
}

#Applying match function to associate each response with a SJR score
q1_SJR<-as.data.frame(apply(q1_responses,2,f_match,SJR,Names))
q2_SJR<-as.data.frame(apply(q2_responses,2,f_match,SJR,Names))

#Calculating the mean SJR scores for each respondent for questions 1 and 2
q1_gender <- cbind(results[3],q1_SJR) %>%
  mutate(mean_SJR=round(rowMeans(q1_SJR,na.rm=TRUE),2))
q2_gender <- cbind(results[3],q2_SJR) %>%
  mutate(mean_SJR=round(rowMeans(q2_SJR,na.rm=TRUE),3))

#Recoding career stage response to either be tenure-track or non-tenure-track
Career_TT <- ifelse(results$Career == "Assistant Professor" |
                      results$Career == "Associate Professor" |
                      results$Career == "Full Professor" |
                      results$Career == "Emeritus Professor", "TT", "Non-TT")

#Calculating the mean SJR scores for each respondent for questions 1 and 2
q1_TT_Career <- cbind(Career_TT, q1_SJR) %>%
  mutate(mean_SJR=round(rowMeans(q1_SJR,na.rm=TRUE),2))
q2_TT_Career <- cbind(Career_TT, q2_SJR) %>%
  mutate(mean_SJR=round(rowMeans(q2_SJR,na.rm=TRUE),2))

#Subsetting question 1 data based on two gender factors (Man/Woman)
q1_Men <- filter(q1_gender, Gender == "Man")  %>% as.data.frame()
q1_Women <- filter(q1_gender, Gender == "Woman")  %>% as.data.frame()

#Subsetting question 1 data based on two career stages (TT/non-TT)
q1_TT <- filter(q1_TT_Career, Career_TT == "TT")  %>% as.data.frame()
q1_non_TT <- filter(q1_TT_Career, Career_TT == "Non-TT")  %>% as.data.frame()

#Subsetting question 2 data based on two gender factors (Man/Woman)
q2_Men <- filter(q2_gender, Gender == "Man") %>% as.data.frame()
q2_Women <- filter(q2_gender, Gender == "Woman")  %>% as.data.frame()

#Subsetting question 2 data based on two career stages (TT/non-TT)
q2_TT <- filter(q2_TT_Career, Career_TT == "TT")  %>% as.data.frame()
q2_non_TT <- filter(q2_TT_Career, Career_TT == "Non-TT")  %>% as.data.frame()

#Performing Mann-Whitney-Wilcoxon test for gender and career stage for questions 1 and 2
wilcox.test(q1_Men$mean_SJR,q1_Women$mean_SJR)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  q1_Men$mean_SJR and q1_Women$mean_SJR
## W = 9951, p-value = 0.2147
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(q1_TT$mean_SJR,q1_non_TT$mean_SJR)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  q1_TT$mean_SJR and q1_non_TT$mean_SJR
## W = 7691, p-value = 0.01423
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(q2_Men$mean_SJR,q2_Women$mean_SJR)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  q2_Men$mean_SJR and q2_Women$mean_SJR
## W = 10448, p-value = 0.003166
## alternative hypothesis: true location shift is not equal to 0
```

```r
wilcox.test(q2_TT$mean_SJR,q2_non_TT$mean_SJR)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  q2_TT$mean_SJR and q2_non_TT$mean_SJR
## W = 7036.5, p-value = 0.005886
## alternative hypothesis: true location shift is not equal to 0
```

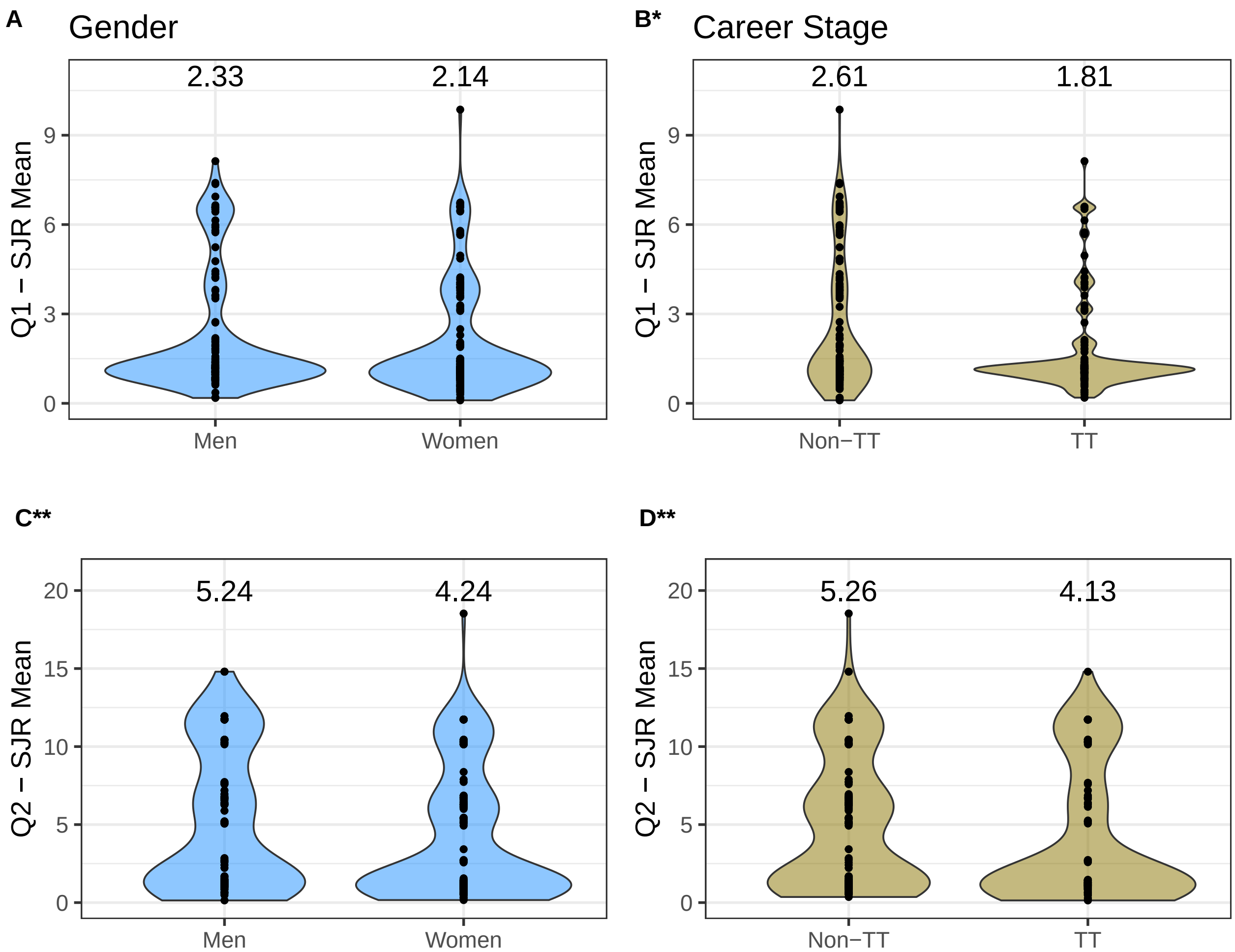
In order to visualize the differences between our groups, we use a series of violin plots. See ?geom_violin for additional information about a violin plot. Broadly, it can be viewed as a display of a continuous distribution and is a blend of a boxplot and a mirrored density plot.

```r
#Combining data from each of the category
#(this is to plot the mean SJR score for each category below)
q1_Men_SJR <- c(q1_Men$Q1_1,q1_Men$Q1_2,q1_Men$Q1_3,
                q1_Men$Q1_4,q1_Men$Q1_5,q1_Men$Q1_6)
q1_Women_SJR <- c(q1_Women$Q1_1,q1_Women$Q1_2,q1_Women$Q1_3,
                  q1_Women$Q1_4,q1_Women$Q1_5,q1_Women$Q1_6)
q1_TT_SJR <- c(q1_TT$Q1_1,q1_TT$Q1_2,q1_TT$Q1_3,
               q1_TT$Q1_4,q1_TT$Q1_5,q1_TT$Q1_6)
q1_non_TT_SJR <- c(q1_non_TT$Q1_1,q1_non_TT$Q1_2,q1_non_TT$Q1_3,
                   q1_non_TT$Q1_4,q1_non_TT$Q1_5,q1_non_TT$Q1_6)
q2_Men_SJR <- c(q2_Men$Q2_1,q2_Men$Q2_2,q2_Men$Q2_3,
                q2_Men$Q2_4,q2_Men$Q2_5,q2_Men$Q2_6)
q2_Women_SJR <- c(q2_Women$Q2_1,q2_Women$Q2_2,q2_Women$Q2_3,
```

```
                  q2_Women$Q2_4,q2_Women$Q2_5,q2_Women$Q2_6)
q2_TT_SJR <- c(q2_TT$Q2_1,q2_TT$Q2_2,q2_TT$Q2_3,
               q2_TT$Q2_4,q2_TT$Q2_5,q2_TT$Q2_6)
q2_non_TT_SJR <- c(q2_non_TT$Q2_1,q2_non_TT$Q2_2,q2_non_TT$Q2_3,
                   q2_non_TT$Q2_4,q2_non_TT$Q2_5,q2_non_TT$Q2_6)

#Violin plot for gender based on question 1 responses
q1_gender_violin<-ggplot(filter(q1_gender, Gender == "Man"| Gender == "Woman"),
                                aes(x=Gender,y=mean_SJR)) +
  geom_violin(fill="dodgerblue",alpha=0.5) +
  geom_point() +
  labs(x="",y="Q1 - SJR Mean",title="Gender") +
  annotate("text",x=c(1,2),y=11,label=round(c(mean(q1_Men_SJR,na.rm=TRUE),
                                        mean(q1_Women_SJR,na.rm=TRUE)),2),
           size=6) +
  ylim(0,11) + theme_bw(base_size = 16) +
  scale_x_discrete(labels=c("Men","Women"))

#Violin plot for gender based on question 2 responses
q2_gender_violin<-ggplot(filter(q2_gender, Gender == "Man" | Gender == "Woman"),
                         aes(x=Gender,y=mean_SJR)) +
  geom_violin(fill="dodgerblue",alpha=0.5) +
  geom_point() +
  labs(x="",y="Q2 - SJR Mean",title="") +
  annotate("text",x=c(1,2),y=20,label=round(c(mean(q2_Men_SJR,na.rm=TRUE),
                                        mean(q2_Women_SJR,na.rm=TRUE)),2),
           size=6) +
  ylim(0,21) + theme_bw(base_size = 16) +
  scale_x_discrete(labels=c("Men","Women"))

#Violin plot for career stage (TT/non-TT) based on question 1 responses
q1_TT_violin<-ggplot(q1_TT_Career, aes(x=Career_TT,y=mean_SJR)) +
  geom_violin(fill="gold4",alpha=0.5) +
  geom_point() +
  labs(x="",y="Q1 - SJR Mean",title="Career Stage") +
  theme_classic(base_size = 22) +
  annotate("text",x=c(1,2),y=11,label=round(c(mean(q1_non_TT_SJR,na.rm=TRUE),
                                        mean(q1_TT_SJR,na.rm=TRUE)),2),size=6) +
  ylim(0,11) + theme_bw(base_size = 16)

#Violin plot for career stage (TT/non-TT) based on question 2 responses
q2_TT_violin<-ggplot(q2_TT_Career, aes(x=Career_TT,y=mean_SJR)) +
  geom_violin(fill="gold4",alpha=0.5) +
  geom_point() +
  labs(x="",y="Q2 - SJR Mean",title="") +
  theme_classic(base_size = 22) +
  annotate("text",x=c(1,2),y=20,label=round(c(mean(q2_non_TT_SJR,na.rm=TRUE),
                                        mean(q2_TT_SJR,na.rm=TRUE)),2),size=6) +
  ylim(0,21) + theme_bw(base_size = 16)
```

The violin plots can be arranged and plotted using the following code:

```r
jpeg("Figure_4.jpg",width=12,height=10,units="in",res=300)
ggarrange(q1_gender_violin, q1_TT_violin, q2_gender_violin, q2_TT_violin,
          ncol=2,nrow=2,
          labels = c("A","B*","C**","D**"))
dev.off()
```



Question 4

What factors do archaeologists say they are considering when they decide where to publish an academic article?

This question is directly address in Question 3 of our survey which asked participants to rank eight different factors that they might consider when publishing an academic article. The analysis of responses from the question is provided below.

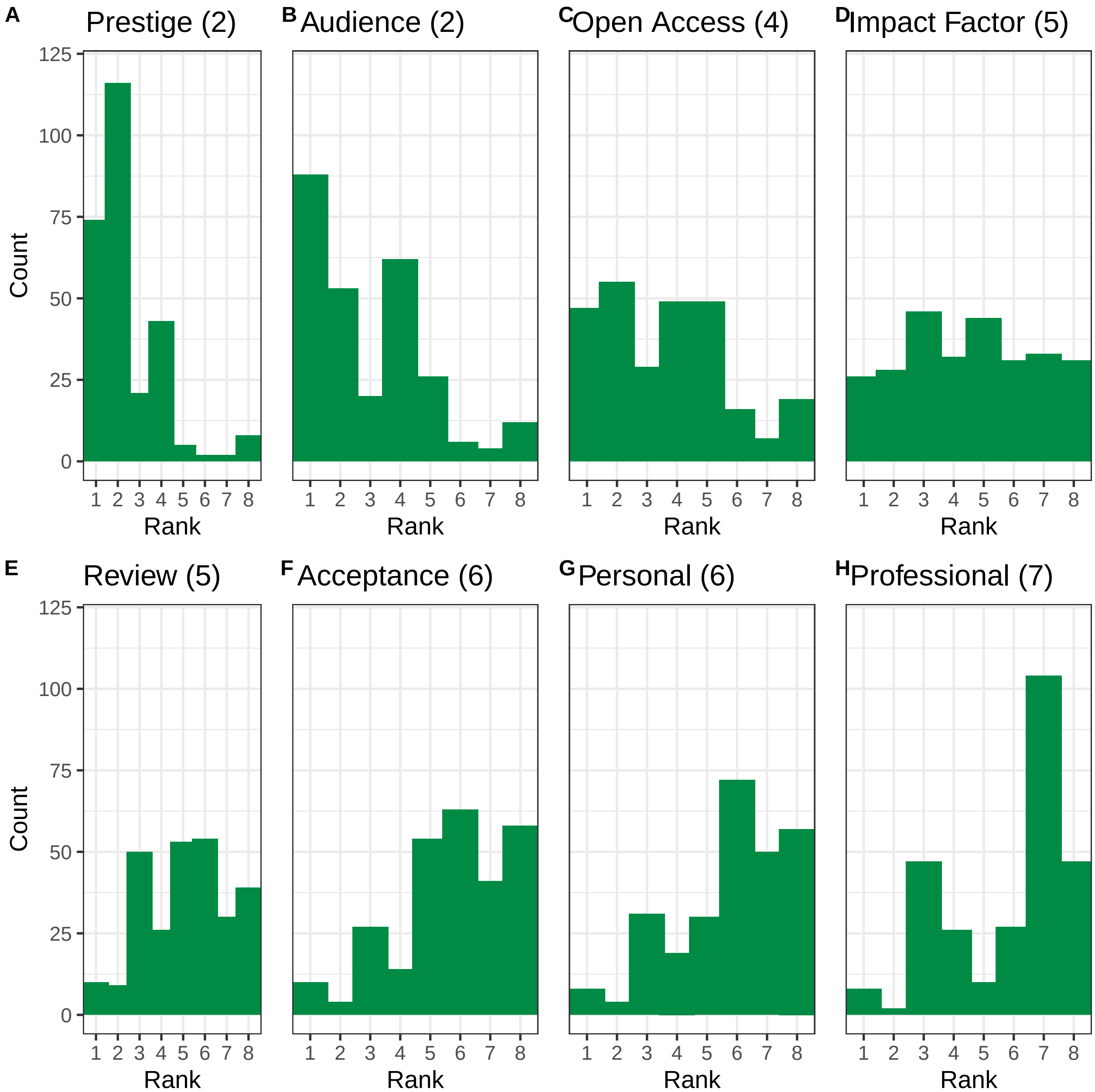
Once we have parsed the data from the survey into usable dataframe, we want to visualize the distribution of rank orders for each factor that was listed (prestige, audience, open access, impact factor, review duration, acceptance rate, professional obligation, personal relationships). These individual histograms can be created using the following:

```r
#Rearranging the data to be in median rank order
q3_sum_t <- as.data.frame(t(q3_sum))
q3_sum_t <- q3_sum_t[-9,]
q3_sum_t$order <- c(1,2,3,4,5,6,7,8)

prestige_hist <- ggplot(q3_sum_t,aes(x=Prestige,
                                     y=reorder(row.names(q3_sum_t),order))) +
  geom_bar(stat="identity",fill="springgreen4",binwidth=0,width=1.2) +
  labs(x="Count",y="Rank",title="Prestige (2)") +
    theme_bw(base_size=16) + coord_flip() +
    xlim(0,120) +
    scale_y_discrete(labels = c("1","2","3","4","5","6","7","8")) +
    theme(plot.title = element_text(hjust = 0.1))

audience_hist <- ggplot(q3_sum_t,aes(x=Audience,
                                     y=reorder(row.names(q3_sum_t),order))) +
  geom_bar(stat="identity",fill="springgreen4",binwidth=0,width=1.2) +
  labs(x="Count",y="Rank",title="Audience (2)") +
    theme_bw(base_size=16) + coord_flip() +
    xlim(0,120) +
    scale_y_discrete(labels = c("1","2","3","4","5","6","7","8")) +
   theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank(),
        plot.title = element_text(hjust = 0.1))

oa_hist <- ggplot(q3_sum_t,aes(x=Open_Access,
                               y=reorder(row.names(q3_sum_t),order))) +
  geom_bar(stat="identity",fill="springgreen4",binwidth=0,width=1.2) +
  labs(x="Count",y="Rank",title="Open Access (4)") +
    theme_bw(base_size=16) + coord_flip() +
  xlim(0,120) +
    scale_y_discrete(labels = c("1","2","3","4","5","6","7","8")) +
   theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank(),
        plot.title = element_text(hjust = 0.1))

impact_hist <- ggplot(q3_sum_t,aes(x=Impact_Factor,
                                   y=reorder(row.names(q3_sum_t),order))) +
  geom_bar(stat="identity",fill="springgreen4",binwidth=0,width=1.2) +
  labs(x="Count",y="Rank",title="Impact Factor (5)") +
    theme_bw(base_size=16) + coord_flip() +
  xlim(0,120) +
    scale_y_discrete(labels = c("1","2","3","4","5","6","7","8")) +
   theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank(),
        plot.title = element_text(hjust = 0.1))

review_hist <- ggplot(q3_sum_t,aes(x=Review_Process,
                                   y=reorder(row.names(q3_sum_t),order))) +
  geom_bar(stat="identity",fill="springgreen4",binwidth=0,width=1.2) +
```

```
  labs(x="Count",y="Rank",title="Review (5)") +
    theme_bw(base_size=16) + coord_flip() +
    xlim(0,120) +
    scale_y_discrete(labels = c("1","2","3","4","5","6","7","8"))

accept_hist <- ggplot(q3_sum_t,aes(x=Accept_Rate,
                                   y=reorder(row.names(q3_sum_t),order))) +
  geom_bar(stat="identity",fill="springgreen4",binwidth=0,width=1.2) +
  labs(x="Count",y="Rank",title="Acceptance (6)") +
    theme_bw(base_size=16) + coord_flip() +
    xlim(0,120) +
    scale_y_discrete(labels = c("1","2","3","4","5","6","7","8")) +
     theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank(),
        plot.title = element_text(hjust = 0.1))

personal_hist <- ggplot(q3_sum_t,aes(x=Personal_Links,
                                     y=reorder(row.names(q3_sum_t),order))) +
  geom_bar(stat="identity",fill="springgreen4",binwidth=0,width=1.2) +
  labs(x="Count",y="Rank",title="Personal (6)") +
    theme_bw(base_size=16) + coord_flip() +
    xlim(0,120) +
    scale_y_discrete(labels = c("1","2","3","4","5","6","7","8")) +
     theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank(),
        plot.title = element_text(hjust = 0.1))

professional_hist <- ggplot(q3_sum_t,aes(x=Prof_Obligation,
                                         y=reorder(row.names(q3_sum_t),order))) +
  geom_bar(stat="identity",fill="springgreen4",binwidth=0,width=1.2) +
  labs(x="Count",y="Rank",title="Professional (7)") +
    theme_bw(base_size=16) + coord_flip() +
    xlim(0,120) +
    scale_y_discrete(labels = c("1","2","3","4","5","6","7","8")) +
     theme(axis.title.y=element_blank(),
        axis.text.y=element_blank(),
        axis.ticks.y=element_blank(),
        plot.title = element_text(hjust = 0.1))
```

The histograms can be arranged and plotted using the following:

```
jpeg("Figure_5.jpg",width=15,height=15,units="in",res=300)
ggarrange(prestige_hist,audience_hist,oa_hist,impact_hist,
          review_hist,accept_hist,personal_hist,professional_hist,ncol=4,nrow=2,
          labels = c("A","B","C","D","E","F","G","H"))
dev.off()
```



SJR Scores by Career Stage

Finally, we were also interested in examining the mean SJR scores for each of our specific career stage category, as opposed to the aggregated categories highlighted above. Here, we use violin plots again to visualize the distribution of mean SJR scores for each survey respondent based on the self-identified career stage.

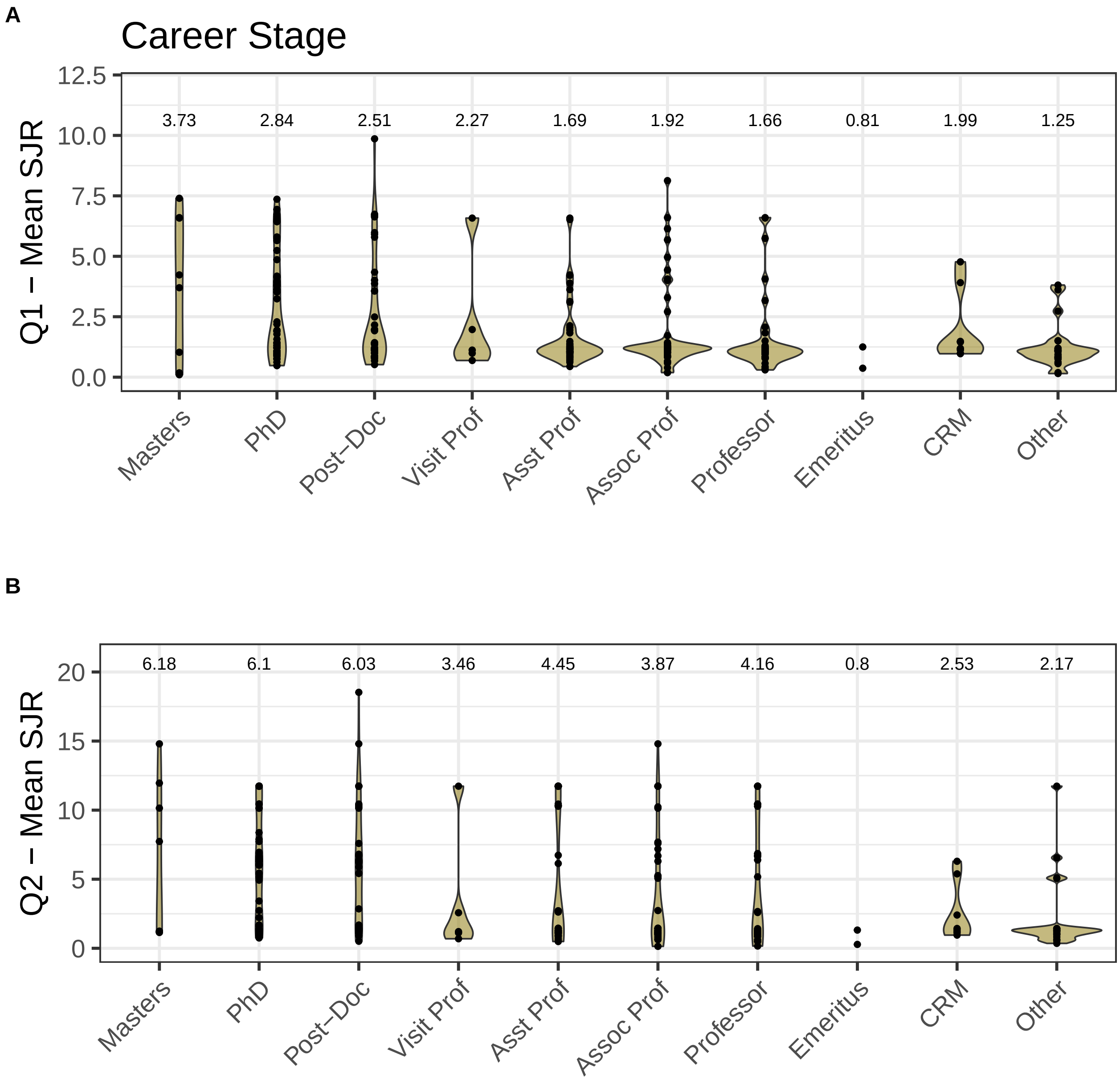
```
#Obtain mean SJR score for each respondent based on career stage for question 1
q1_career <- cbind(results[8],q1_SJR) %>%
  mutate(mean_SJR=round(rowMeans(q1_SJR,na.rm=TRUE),2))

#Recoding career stages with shorter names for better plotting
```

```r
q1_career$Position_Short <-
    ifelse(q1_career$Career == "Assistant Professor","Asst Prof",
    ifelse(q1_career$Career == "Associate Professor", "Assoc Prof",
    ifelse(q1_career$Career == "Full Professor", "Professor",
    ifelse(q1_career$Career == "Emeritus Professor", "Emeritus",
    ifelse(q1_career$Career == "Master's Student","Masters",
    ifelse(q1_career$Career=="PhD Student","PhD",
    ifelse(q1_career$Career=="Post-Doc","Post-Doc",
    ifelse(q1_career$Career=="Visiting Professor", "Visit Prof",
    ifelse(q1_career$Career=="CRM/Contract Archaeologist","CRM","Other")))))))))

#Adding in shorten names and joining with previous and also
#reordering x-axis based chronological order of career stage
q1_career_axis <- data.frame(Position_Short=c("PhD", "Asst Prof", "Assoc Prof",
                                              "Post-Doc","Professor","Other",
                                              "CRM","Masters","Visit Prof",
                                              "Emeritus"),
                             car_x_pos=c(2,5,6,3,7,10,9,1,4,8))
q1_career_order<- left_join(q1_career, q1_career_axis, by="Position_Short")

#Obtain mean SJR score for each respondent based on career stage for question 2
q2_career <- cbind(results[8],q2_SJR) %>%
  mutate(mean_SJR=round(rowMeans(q2_SJR,na.rm=TRUE),3))

#Recoding career stages with shorter names for better plotting for question 2
q2_career$Position_Short <-
    ifelse(q1_career$Career == "Assistant Professor","Asst Prof",
    ifelse(q1_career$Career == "Associate Professor", "Assoc Prof",
    ifelse(q1_career$Career == "Full Professor", "Professor",
    ifelse(q1_career$Career == "Emeritus Professor", "Emeritus",
    ifelse(q1_career$Career == "Master's Student","Masters",
    ifelse(q1_career$Career=="PhD Student","PhD",
    ifelse(q1_career$Career=="Post-Doc","Post-Doc",
    ifelse(q1_career$Career=="Visiting Professor", "Visit Prof",
    ifelse(q1_career$Career=="CRM/Contract Archaeologist","CRM","Other")))))))))

#Adding in shorten names and joining with previous and also
#reordering x-axis based chronological order of career stage for question 2
q2_career_axis <- data.frame(Position_Short=c("PhD", "Asst Prof", "Assoc Prof",
                                              "Post-Doc","Professor","Other",
                                              "CRM","Masters","Visit Prof",
                                              "Emeritus"),
                             car_x_pos=c(2,5,6,3,7,10,9,1,4,8))
q2_career_order<- left_join(q2_career, q2_career_axis, by="Position_Short")

#Joining together ordered career stages with mean SJRs for respondents - Q1
q1_career_means <- q1_career_order %>% group_by(Career) %>%
  summarise(mean=mean(mean_SJR,na.rm=TRUE)) %>%
  mutate(Position_Short=c("Asst Prof","Assoc Prof","CRM", "Emeritus",
                          "Professor", "Masters","Other","PhD",
                          "Post-Doc","Visit Prof")) %>%
  left_join(career_axis, by="Position_Short") %>% arrange(car_x_pos)
```

```r
#Joining together ordered career stages with mean SJRs for respondents - Q2
q2_career_means <- q2_career_order %>% group_by(Career) %>%
  summarise(mean=mean(mean_SJR,na.rm=TRUE)) %>%
  mutate(Position_Short=c("Asst Prof","Assoc Prof","CRM", "Emeritus",
                          "Professor", "Masters","Other","PhD",
                          "Post-Doc","Visit Prof")) %>%
  left_join(career_axis, by="Position_Short") %>% arrange(car_x_pos)

#Histogram for question 1
q1_career_hist<-ggplot(q1_career_order,aes(x=reorder(Position_Short,car_x_pos),
                                          y=mean_SJR)) +
  geom_violin(fill="gold4",alpha=0.5) +
  geom_point() +
  labs(x="",y="Q1 - Mean SJR",title="Career Stage") +
  annotate("text",x=c(1,2,3,4,5,6,7,8,9,10),y=10,
           label = round(q1_career_means$mean,2),
            vjust=-0.75,
            color = "black",
            size = 4) +
  theme_bw(base_size=20) +
  ylim(0,12) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.margin = unit(c(0.5,0.5,0.5,0.5),"cm"))

#Histogram for question 2
q2_career_hist<-ggplot(q2_career_order,aes(x=reorder(Position_Short,car_x_pos),
                                          y=mean_SJR)) +
  geom_violin(fill="gold4",alpha=0.5) +
  geom_point() +
  labs(x="",y="Q2 - Mean SJR",title="") +
  annotate("text",x=c(1,2,3,4,5,6,7,8,9,10),y=19.5,
           label = round(q2_career_means$mean,2),
            vjust=-0.75,
            color = "black",
            size = 4) +
  theme_bw(base_size=20) +
  ylim(0,21) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1),
        plot.margin = unit(c(0.5,0.5,0.5,0.5),"cm"))
```

And the associated plot, can be created with the following:

```r
jpeg("Figure_6.jpg",width = 10, height = 10, units="in", res = 300)
ggarrange(q1_career_hist,q2_career_hist,nrow=2,ncol=1,labels=c("A","B"))
dev.off()
```


A
Career Stage
Q1 – Mean SJR
12.5
10.0
7.5
5.0
2.5
0.0
3.73
2.84
2.51
2.27
1.69
1.92
1.66
0.81
1.99
1.25
Masters
PhD
Post–Doc
Visit Prof
Asst Prof
Assoc Prof
Professor
Emeritus
CRM
Other
B
Q2 – Mean SJR
20
15
10
5
0
6.18
6.1
6.03
3.46
4.45
3.87
4.16
0.8
2.53
2.17
Masters
PhD
Post–Doc
Visit Prof
Asst Prof
Assoc Prof
Professor
Emeritus
CRM
Other