**R: Introduction Workshop**

**Instructor**

Christina Maimone

[christina.maimone@northwestern.edu](mailto:christina.maimone@northwestern.edu)

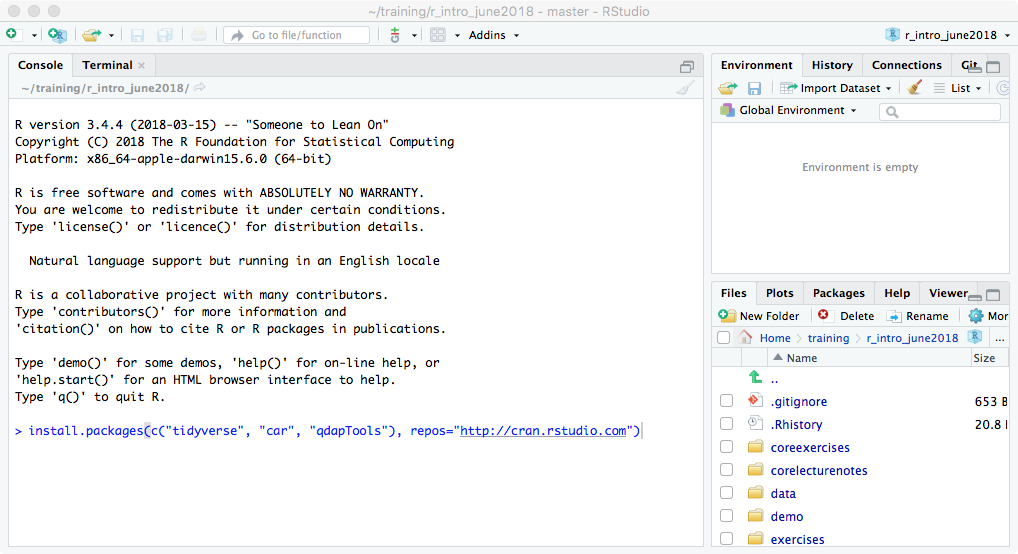
https://www.it.northwestern.edu/research/consultation/data-services.html

**Workshop Materials**

https://github.com/nuitrcs/r-intro-summer2019/

1. Download materials
2. Open Project in RStudio (instructions at repository above)
3. In RStudio, in R console, install packages (ask neighbors or flag down helpers)
   1. If asked whether you want to compile packages (or something similar) say NO

install.packages(c("tidyverse", "car", "qdapTools"), repos="http://cran.rstudio.com")



**Northwestern IT Research Computing Services**

<https://www.it.northwestern.edu/research/consultation/data-services.html>

Example Analysis Script

library(lubridate)

library(qdapTools)

safi <- read.csv("data/SAFI\_clean.csv", na = c("", "NULL", "NA"), stringsAsFactors = FALSE)

View(safi)

# fix some variables

safi$village <- factor(safi$village) # categorical

safi$respondent\_wall\_type <- factor(trimws(safi$respondent\_wall\_type)) # categorical

safi$affect\_conflicts <- factor(safi$affect\_conflicts, ordered=TRUE,

levels=c("never","once","more\_once","frequently"))

safi$interview\_date <- ymd\_hms(safi$interview\_date) # date

safi$memb\_assoc <- ifelse(is.na(safi$memb\_assoc), NA,

ifelse(safi$memb\_assoc == "yes", TRUE, FALSE)) # yes/no to T/F

# convert delimited text field to indicator variables

month\_indicators <- mtabulate(strsplit(safi$months\_lack\_food, ";"))

names(month\_indicators) <- substr(names(month\_indicators), 0, 3) # shorten column names

month\_indicators <- month\_indicators[,month.abb] # reorder columns

safi <- cbind(safi, month\_indicators)

safi$months\_lack\_food\_count <- apply(month\_indicators, 1, sum) # count months lacking food

# in which month are respondents most likely to lack food?

apply(month\_indicators, 2, sum)

which.max(apply(month\_indicators, 2, sum))

table(safi$village) # how many people in each village?

# do respondents in Chirodzo lack food in more months on average than respondents in Ruaca?

tapply(safi$months\_lack\_food\_count, safi$village, mean)

t.test(safi$months\_lack\_food\_count[safi$village == "Chirodzo"],

safi$months\_lack\_food\_count[safi$village == "Ruaca"])

# generalize: anova of village effect on months without food

a1 <- aov(months\_lack\_food\_count ~ -1+village, data = safi) # save anova results

a1 # print results

summary(a1) # summary gives you better formatted results

print(model.tables(a1)) # print the means

summary.lm(a1) # see the regression (lm) that was run for the anova

# plot distribution of family size

barplot(table(safi$no\_membrs)) # gaps aren't shown, so...

barplot(table(factor(safi$no\_membrs, levels=1:max(safi$no\_membrs))),

main="Family Size", xlab="Number of Members",

ylab="Number of Families")

hist(safi$years\_liv, col="gray") # distribution of years\_liv

hist(safi$years\_liv, col="gray", breaks=20) # different number of bars

boxplot(no\_membrs~village, data=safi) # boxplot for number of family members by village

# scatter plots of number of family members and months of lack of food

plot(y=safi$months\_lack\_food\_count, x=safi$no\_membrs)

plot(y=jitter(safi$months\_lack\_food\_count), x=jitter(safi$no\_membrs),

col=rgb(.1, 0, 1, alpha=.3), pch=16)

plot(y=safi$months\_lack\_food\_count, x=safi$no\_membrs,

col=rgb(.1, 0, 1, alpha=.3), pch=16, cex=3)

smoothScatter(y=safi$months\_lack\_food\_count, x=safi$no\_membrs)