

Dustin Tracy
RE2P
April 4, 2014

R Workshop

"R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS."

<http://www.r-project.org/> Download it and packages from: <http://cran.wustl.edu/>

There are also editors, e.g. <https://www.rstudio.com/>

```
setwd("C:/Users/Dustin/Documents/GSU/R")
getwd()

source("C:\\Users\\Dustin\\Documents\\GSU\\R\\R_Workshop.r")

rm(list=ls()) #clears workspace
#I'm a comment line

? rm
help(ls)

library("stargazer", lib.loc="C:/Users/Dustin/Documents/R/win-library/3.0")
# library("XLConnect", lib.loc="C:/Users/Dustin/Documents/R/win-library/3.0")
library("ggplot2", lib.loc="C:/Users/Dustin/Documents/R/win-library/3.0")
library("grid", lib.loc="C:/Program Files/R/R-3.0.3/library")
library("gridExtra", lib.loc="C:/Users/Dustin/Documents/R/win-library/3.0")
library("plm", lib.loc="C:/Users/Dustin/Documents/R/win-library/3.0")
```

R objects are scalars, vectors, matrices

Scalar:

```
Inflate = 1.03
Inflate <- 1.03 # relic
```

Basic Math

```
inf_20 = Inflate^20
spike = 1.1^5
inf_25 = inf_20 * spike
reg = 800
OT = 134
wage = reg + OT
wpm = 365.25/(7*12)
wpm #to view it
```

Vector:

```
new_obs = c(952, 40, 107, 42, 16, 7, 13, 33, 0, 0, 1, 1, 3, 0, 12, 12)
length(new_obs) #returns length
```

```
new_obs = c(new_obs, log(952))
exp(log(952))
length(new_obs)
new_obs #to view it
```

Importing data / Matrices

```
dir()#tells you what is in directory #abbreviation
file = dir()[7]
file
rm(wage_data) # same as above
file = "Wage2.csv"
wage_data = read.csv("Wage2.csv")
wage_data = read.csv("C:/Users/Dustin/Documents/GSU/R/Wage2.csv")
wage_data = read.csv(file)

#paste()

# read.table(something.txt)

ls()#tells you what objects there are in workspace
```

Examine, View, Edit

```
View(wage_data)
edit(wage_data)

dim(wage_data)
dim(wage_data)[2] #same as length
summary(wage_data$educ)

Wage = wage_data$wage #Matrices are made up of vectors
Hours = wage_data[,2]
rate = Wage/(Hours*wpm) #Math operations work on vectors (and matrices)

new_data = cbind(wage_data, rate)

new_data = rbind(new_data, new_obs)
purge_data = new_data[!is.na(new_data[,16]),] #note != not

obs = dim(purge_data)[1]
```

Loops and Flow control

```
for(i in 1:obs){
    if(is.na(purge_data[i,14])){purge_data[i,14]=0}
}

grad_sch = rep(0,742) #creates a column of zeros
col_grad = rep(0,length(purge_data[,1]))
```

```

HS_grad = rep(0,dim(purge_data)[1])
drop_out = rep(0,obs)

k=obs
while(k>=1){
    if(purge_data$educ[k]>=18) {grad_sch[k]=1}
    else if(purge_data$educ[k]>=16) {col_grad[k]=1}
    else if(purge_data$educ[k]>=12) {HS_grad[k]=1}
    else {drop_out[k]=1}
    k=k-1
} #created dummies

ed_lvl = rep(0,length(purge_data[,1]))

k=obs
while(k>=1){
    if(purge_data$educ[k]>=18) {ed_lvl[k]="Grad School"}
    else if(purge_data$educ[k]>=16) {ed_lvl[k]="College"}
    else if(purge_data$educ[k]>=12) {ed_lvl[k]="High School"}
    else {ed_lvl[k]="Below"}
    k=k-1
}

# repeat and switch other options for loops

dummy_data = cbind(purge_data, grad_sch, col_grad, HS_grad, drop_out, ed_lvl)

write.csv (dummy_data, "dummy_data.csv", row.names=FALSE)

plot(purge_data$IQ, purge_data$wage, ylab="Wage", xlab="IQ")
dev.copy(pdf, file="purged.pdf", height=7, width=7)
dev.off()

mar = c("unmarried", "married")

#plot <- ggplot(dummy_data, aes(IQ, wage)) # y= x= not nec

jpeg('all.jpg')
plot <- ggplot(dummy_data, aes(y=wage, x=IQ)) + geom_point(aes(colour = ed_lvl,
shape = factor(married)))
plot + labs(colour = "Education Level") + labs(shape = "Marital Status") +
scale_shape(labels = mar)
dev.off()

#plot + scale_shape_identity("", c("married" = 1,"unmarried" = 0))

plot_S <- ggplot(data=subset(dummy_data, south = 1), aes(y=wage, x=IQ)) +
geom_point(aes(colour = ed_lvl, shape = factor(married)))
plot_S + labs(title = "South") + labs(colour = "Education Level") + labs(shape
= "Marital Status") + scale_shape(labels = mar)

```

```

dev.copy(png, file="south.png")
dev.off()

plot_NS <- ggplot(data=subset(dummy_data, south = 0), aes(y=wage, x=IQ)) +
geom_point(aes(colour = ed_lvl, shape = factor(married)))
plot_NS + labs(title = "Non-South") + labs(colour = "Education Level") +
labs(shape = "Marital Status") + scale_shape(labels = mar)
dev.copy(png, file="not.png")
dev.off()

# grid.arrange(plot_S, plot_NS)

```

Regressions and Exporting

```

ols1<-lm(wage ~ hours + IQ + KWW + educ, data=dummy_data)
summary(ols1)

ols2<-lm(wage ~ hours + IQ + KWW + educ + exper + tenure, data=dummy_data)
summary(ols2)

ols3<-lm(wage ~ hours + IQ + KWW + educ + exper + tenure + meduc + feduc,
data=dummy_data)
summary(ols3)

stargazer(ols1,ols2,ols3, summary=TRUE, type="text", out="3regs.tex")
# LaTeX code

stargazer(ols1,ols2,ols3, summary=TRUE, type="text", out="3regs.txt")
# text, can be opened in excel; data to columns, fixed width, drag column
widths, - -> () other number options

stargazer(ols1,ols2,ols3, summary=TRUE, type="text", out="3regs.html")
# html, can be opened in a browser, pasted into excel, - -> () other number
options

coll<-summary(lm(wage ~ hours + IQ + KWW + educ, data=dummy_data))$coeff
#writeWorksheetToFile(file='test.xlsx',data=coll,sheet='test2')
write.csv(coll, "simple.csv")

```

Panel Data

```

data("Grunfeld", package="plm")

grun.fe <- plm(inv~value+capital,data=Grunfeld,model="within")
grun.re <- plm(inv~value+capital,data=Grunfeld,model="random")

summary(grun.re)
summary(grun.fe)

stargazer(grun.re,grun.re, summary=TRUE, type="text", out="panel.tex")

```