

GR4058 Assignment 1

Due: Tuesday, September 26, 2017 by 6PM

Instructions

Create a RMarkdown file that contains the answers to the following questions, remembering that even correct code can result in points deducted if it exhibits poor or inconsistent style. Verify that you can knit the RMarkdown file to an HTML file or PDF file. Upload both the RMarkdown file and the resulting HTML or PDF file to the Assignments section of Canvas when you are finished.

Work on this problem set by yourself, but you can ask questions on CampusWire. Just remember to click on the options that says your question will be visible to “Instructors and TAs only”.

Data structures

In what ways does a vector differ from a data.frame in R? Use R code to create one of each.

Style

Rewrite the following code so that it is consistent with Hadley Wickham’s R style guide.

```
a <- c( 5, -2, 3, -4, 1, NA )
b <- a * - 1
b[ b>0 ]

num = 1
mycondition <- round (runif (1,0,1,))
if( mycondition ) {
  num <- num + 1
  print(paste("num =" , num , sep=" "))
} else {
  print("false")
}

# make an integer sequence from 1 to 50
y <- 1:50

# test if an number is even
even <- y%%2 == 0

# subset by the test above
y = y [even]

x <- matrix( c( 23, 34, 35, 6, 87, 39, 21, 14, 99 ) ,nrow = 3)
df <- as.data.frame (x)
names( df ) <- c("percentage_score_on_reading_test",
  "percentage.score.on.math.test",
  "percentage-score-on-writing-test")

mean <- function (x)
{
  sum(x)/length(x)
```

```
}
```

Subsetting a data.frame

In the RStudio console, call

```
View(state.x77)
```

to look at a spreadsheet-like representation of these data.

Then write R code that extracts the data for Connecticut on the variables Population and Life Exp in three different ways.

Manipulating data

In the RStudio console, type

```
help(airquality)
```

to see information on a `data.frame` regarding the air quality in New York. Then:

Replace the missing values on the `Solar.R` variable with the average value of `Solar.R` among non-missing values. Hint: see the `is.na()` function.

American National Election Study

Execute the following once, *outside* your RMarkdown file

```
devtools::install_github("kosukeimai/qss-package", build_vignettes = TRUE)
```

although you may have to first install the **devtools** package from CRAN via

```
install.packages("devtools")
```

Then, add the following line to a R chunk *inside* your RMarkdown file

```
data("turnout", package = "qss")
```

which brings a `data.frame` called `turnout` into R's memory whose variables are described under `help("turnout", package = "qss")` in R. The focus here is on comparing the estimated turnout rate from the American National Election Study (ANES) — which is a survey — to the vote totals and population data.

Calculate the turnout rate based on the voting age population or VAP. Note that for this data set, we must add the total number of eligible overseas voters since the VAP variable does not include these individuals in the count. Next, calculate the turnout rate using the voting eligible population or VEP. What difference do you observe?