

R is also for Filipino Researchers

Joseph S. Tabadero, Jr.

October 28, 2017

Why is “R also for Filipino Researchers”?

I have no experience with coding so R frightens me.

Why use Rstudio with R?

Script file and Rmarkdown

Importing Data into R is easy and can be done in many ways

```
## + ggplot2 2.2.1.9000      Date: 2017-10-31
## + tibble  1.3.4           R: 3.4.2
## + tidyr   0.7.2           OS: elementary OS 0.4.1 Lo
## + readr    1.1.1          GUI: X11
## + purrr    0.2.4          Locale: en_US.UTF-8
## + dplyr    0.7.4          TZ: Asia/Manila
## + stringr  1.2.0
## + forcats  0.2.0
```

```
## Conflicts
```

```
## * filter(), from dplyr, masks stats::filter()
## * lag(),    from dplyr, masks stats::lag()
```

Introducing the `mtcars` data set

What is the `mtcars` data set?

How can we load the `mtcars` data set?

How can we inspect the `mtcars` data set?

How can we visualize the `mtcars` data set?

How can we manipulate the `mtcars` data set?

How can we save the `mtcars` data set?

How can we use the `mtcars` data set?

How can we use the `mtcars` data set?

How can we use the `mtcars` data set?

How can we use the `mtcars` data set?

Some descriptive statistics

A research question: Is there a difference in millege for automatic and manual cars?

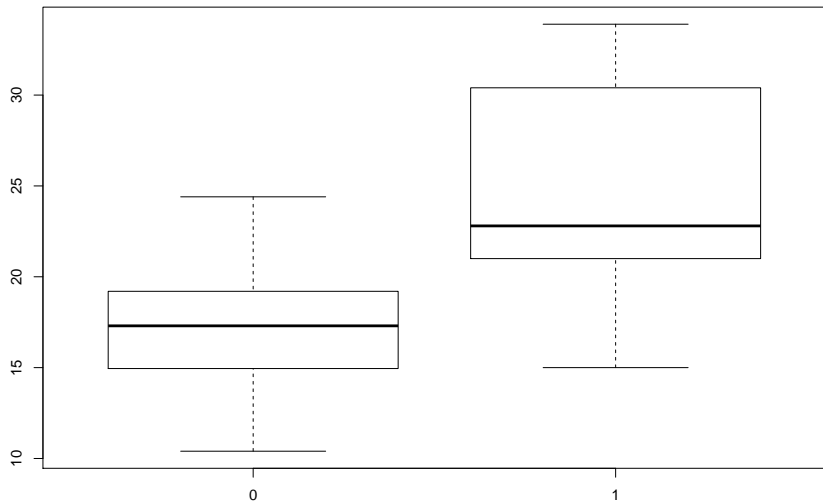
```
## # A tibble: 2 x 4
##       am mean_mpg var_mpg      n
##   <dbl>   <dbl>   <dbl> <int>
## 1     0    17.15    14.70    19
## 2     1    24.39    38.03    13

##
##    0    1
## 19 13

##   Group.1      x
## 1       0 17.15
## 2       1 24.39

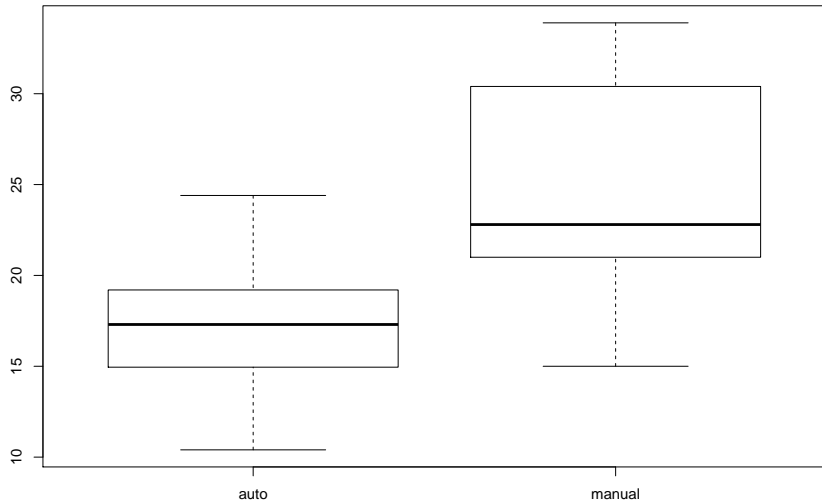
##   Group.1      x
## 1       0 14.70
## 2       1 38.03
```

A research question: Is there a difference in milleage for automatic and manual cars?



Changing the labels of a plot; creating a new variable in a data set (`data.frame`)

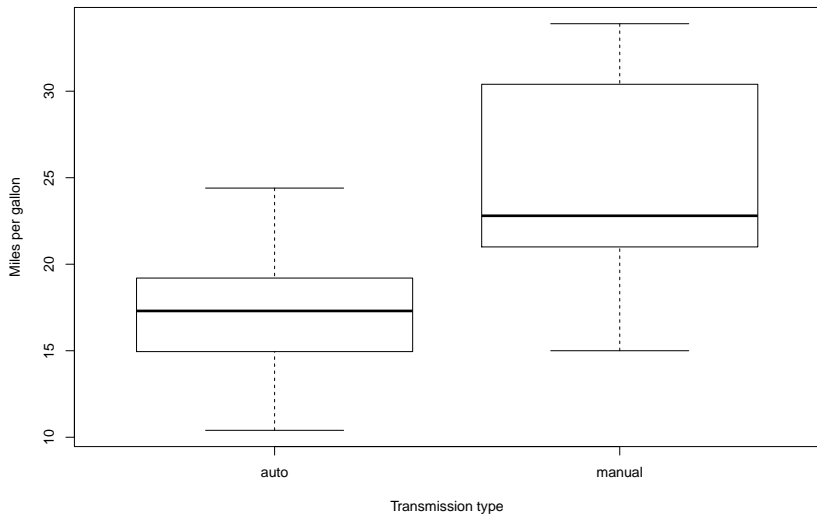
Let us put some labels for the levels of `am`.



Changing the x and y labels and putting a title

Let us put some labels for the levels of `am`.

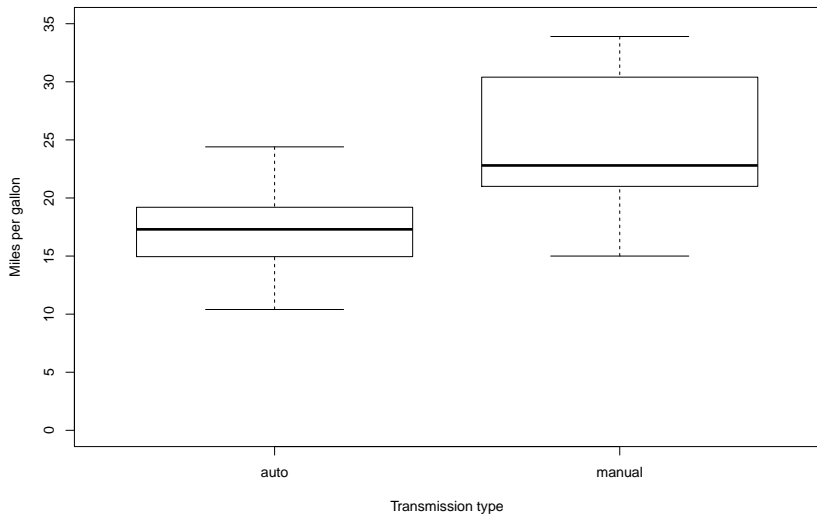
Boxplots of miles per gallon according to transmission type



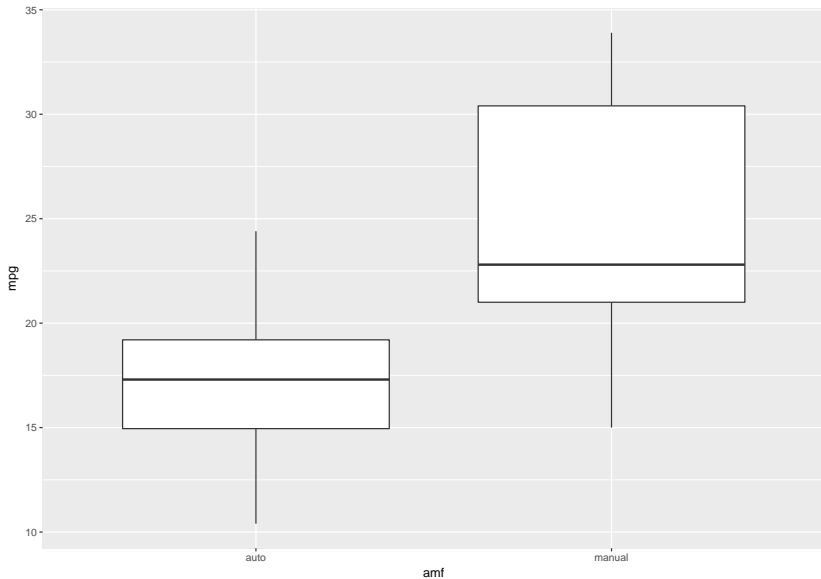
Changing the range of values in the y-axis

Let us put some labels for the levels of `am`.

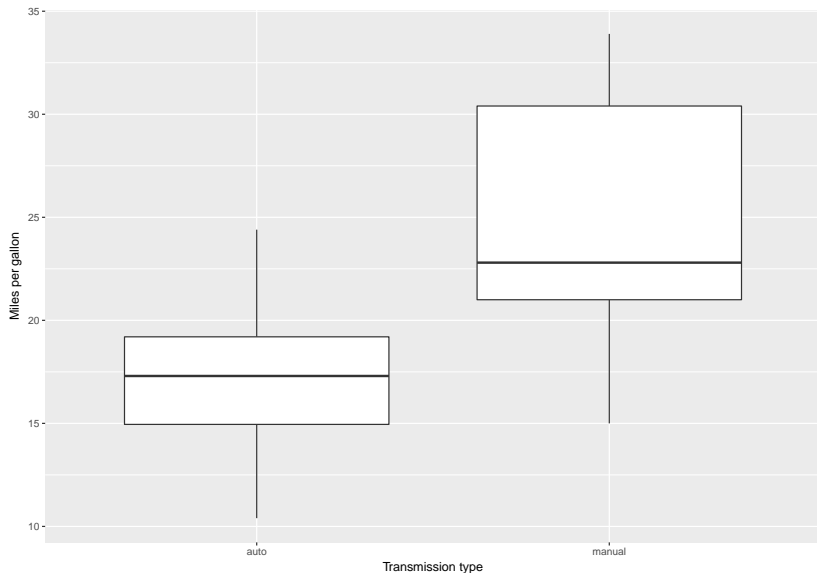
Boxplots of miles per gallon according to transmission type



Plotting with ggplot2



ggplot2 uses the language of graphics



A review of t test

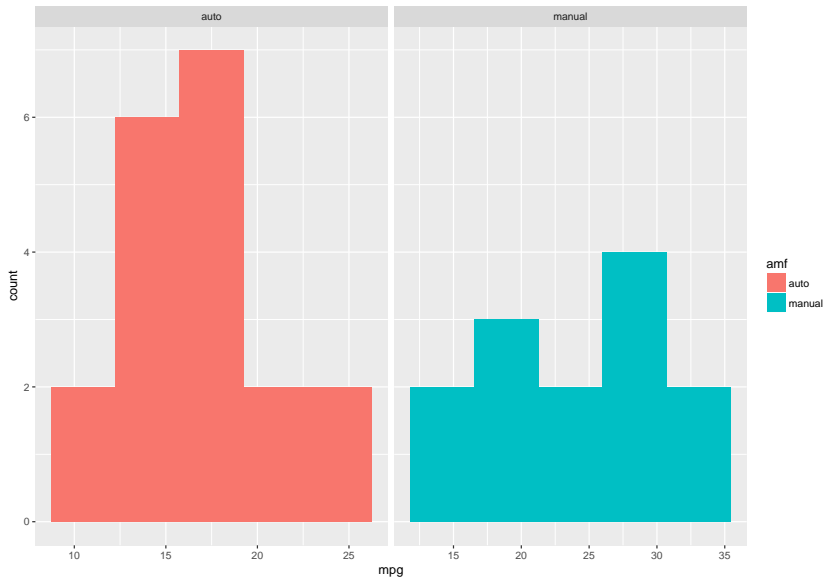
What are the assumptions of the independent samples t test?

1. Dependent variable should be measured on a continuous scale (interval or ratio level)
2. Independent variable consist of two categorical, independent groups
3. Observations are independent of each other
4. No significant outliers
5. Dependent variable should be (approximately) normally distributed for each group of the independent variable
6. Variances should be homogenous

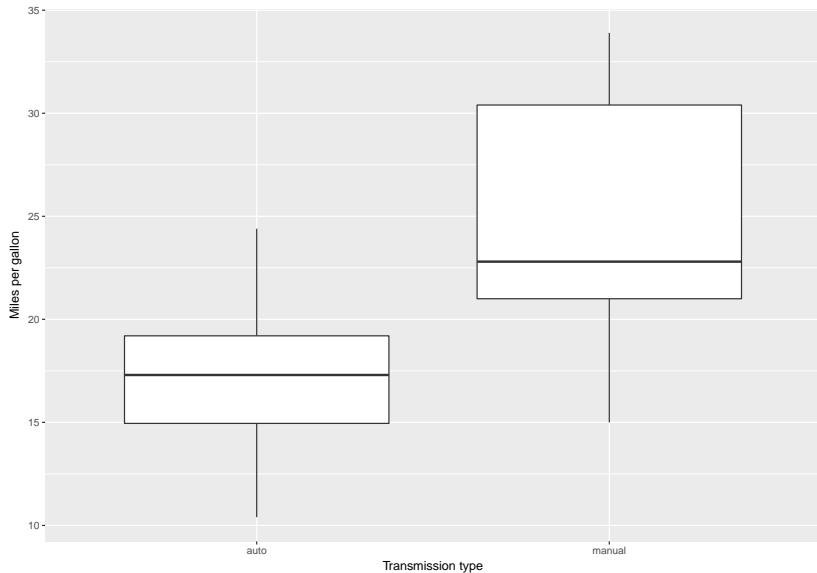
Applying the assumptions

1. What is the independent variable? What is the independent variable?
2. Is the dependent variable measured on a continuous scale?
3. Does the independent variable consist of two categorical, independent groups?
4. Are observations independent of each other?
5. Are there no significant outliers?
6. Is the dependent variable normally distributed for each group of the independent variable?
7. Are the variances homogenous?

Normality for each group of the independent variable



Homogeneity of variance



Homogeneity of variance (cont...)

```
##
```

```
## F test to compare two variances
```

```
##
```

```
## data: mpg by amf
```

```
## F = 0.39, num df = 18, denom df = 12, p-value = 0.07
```

```
## alternative hypothesis: true ratio of variances is not e
```

```
## 95 percent confidence interval:
```

```
## 0.1244 1.0703
```

```
## sample estimates:
```

```
## ratio of variances
```

```
## 0.3866
```

t test results

```
##  
##  Welch Two Sample t-test  
##  
## data:  mpg by amf  
## t = -3.8, df = 18, p-value = 0.001  
## alternative hypothesis: true difference in means is not  
## 95 percent confidence interval:  
##  -11.28  -3.21  
## sample estimates:  
##      mean in group auto mean in group manual  
##                17.15                24.39
```

How to use t.test?

```
?t.test
```

How about independent samples t test?

```
##  
## Two Sample t-test  
##  
## data: mpg by amf  
## t = -4.1, df = 30, p-value = 3e-04  
## alternative hypothesis: true difference in means is not  
## 95 percent confidence interval:  
## -10.848 -3.642  
## sample estimates:  
## mean in group auto mean in group manual  
## 17.15 24.39
```

Confidence intervals

```
## [1] -11.28 -3.21  
## attr(,"conf.level")  
## [1] 0.95
```

```
## [1] -10.848 -3.642  
## attr(,"conf.level")  
## [1] 0.95
```


Non-parametric alternative

```
## Warning in wilcox.test.default(x = c(21.4, 18.7, 18.1, 1
## cannot compute exact p-value with ties

## Warning in wilcox.test.default(x = c(21.4, 18.7, 18.1, 1
## cannot compute exact confidence intervals with ties

##
## Wilcoxon rank sum test with continuity correction
##
## data:  mpg by amf
## W = 42, p-value = 0.002
## alternative hypothesis: true location shift is not equal
## 95 percent confidence interval:
##  -11.7  -2.9
## sample estimates:
## difference in location
##                -6.8
```