

```
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':  
##  
##   filter, lag
```

```
## The following objects are masked from 'package:base':  
##  
##   intersect, setdiff, setequal, union
```

```
library(ggplot2)  
library(car)
```

```
## Loading required package: carData
```

```
##  
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':  
##  
##   recode
```

```
library(ggpubr)  
library(tidyverse)
```

```
## — Attaching packages ————— tidyverse 1.3.1 —
```

```
## ✓ tibble 3.1.7    ✓ purrr  0.3.4  
## ✓ tidyr  1.2.0    ✓ stringr 1.4.0  
## ✓ readr  2.1.2    ✓ forcats 0.5.1
```

```
## — Conflicts ————— tidyverse_conflicts() —  
## X dplyr::filter() masks stats::filter()  
## X dplyr::lag()    masks stats::lag()  
## X car::recode()   masks dplyr::recode()  
## X purrr::some()   masks car::some()
```

```
library(rstatix)
```

```
##
## Attaching package: 'rstatix'
```

```
## The following object is masked from 'package:stats':
##
## filter
```

```
library(ggstatsplot)
```

```
## Registered S3 method overwritten by 'parameters':
## method from
## format.parameters_distribution datawizard
```

```
## You can cite this package as:
## Patil, I. (2021). Visualizations with statistical details: The 'ggstatsplot' approach.
## Journal of Open Source Software, 6(61), 3167, doi:10.21105/joss.03167
```

```
setwd("C:/Users/Laura/Documents/code/MA-541")
df <- read.csv("Crime_R.csv")
# split data into year 0 and year + 10
dim(df)
```

```
## [1] 47 27
```

```
names(df)
```

```
## [1] "CrimeRate" "Youth" "Southern"
## [4] "Education" "ExpenditureYear0" "LabourForce"
## [7] "Males" "MoreMales" "StateSize"
## [10] "YouthUnemployment" "MatureUnemployment" "HighYouthUnemploy"
## [13] "Wage" "BelowWage" "CrimeRate10"
## [16] "Youth10" "Education10" "ExpenditureYear10"
## [19] "LabourForce10" "Males10" "MoreMales10"
## [22] "StateSize10" "YouthUnemploy10" "MatureUnemploy10"
## [25] "HighYouthUnemploy10" "Wage10" "BelowWage10"
```

```
head(df,2)
```

```
##   CrimeRate Youth Southern Education ExpenditureYear0 LabourForce Males
## 1      45.5   135         0      12.4                69        540   965
## 2      52.3   140         0      10.9                55        535  1045
##   MoreMales StateSize YouthUnemployment MatureUnemployment HighYouthUnemploy
## 1         0         6                80                22                1
## 2         1         6               135                40                1
##   Wage BelowWage CrimeRate10 Youth10 Education10 ExpenditureYear10
## 1   564        139        26.5   135        12.5                71
## 2   453        200        35.9   135        10.9                54
##   LabourForce10 Males10 MoreMales10 StateSize10 YouthUnemploy10
## 1         564       974         0         6                82
## 2         540      1039         1         7               138
##   MatureUnemploy10 HighYouthUnemploy10 Wage10 BelowWage10
## 1                20                1   632        142
## 2                39                1   521        210
```

```
tail(df,2)
```

```
##   CrimeRate Youth Southern Education ExpenditureYear0 LabourForce Males
## 46      157.7   136         0      15.1                149        577   994
## 47      161.8   131         0      13.2                160        631  1071
##   MoreMales StateSize YouthUnemployment MatureUnemployment HighYouthUnemploy
## 46         0        157                102                39                0
## 47         1         3                102                41                0
##   Wage BelowWage CrimeRate10 Youth10 Education10 ExpenditureYear10
## 46   673        167        177.2   140        15.2                141
## 47   674        152        178.2   132        13.2                143
##   LabourForce10 Males10 MoreMales10 StateSize10 YouthUnemploy10
## 46         578       995         0        160                110
## 47         632      1058         1         4                100
##   MatureUnemploy10 HighYouthUnemploy10 Wage10 BelowWage10
## 46                40                0   739        169
## 47                40                0   748        150
```

```
df0 <- df %>%
  select(-ends_with('10'))
df10 <- df %>%
  select(ends_with('10'))

#str(df)
```

```
df1 <- df[,c("Southern", "Males")]
df2 <- df[,c("Southern", "Males10")]
df2$Southern <- ifelse(df2$Southern == 0, 3, 4)
names(df2)[2] <- "Males"

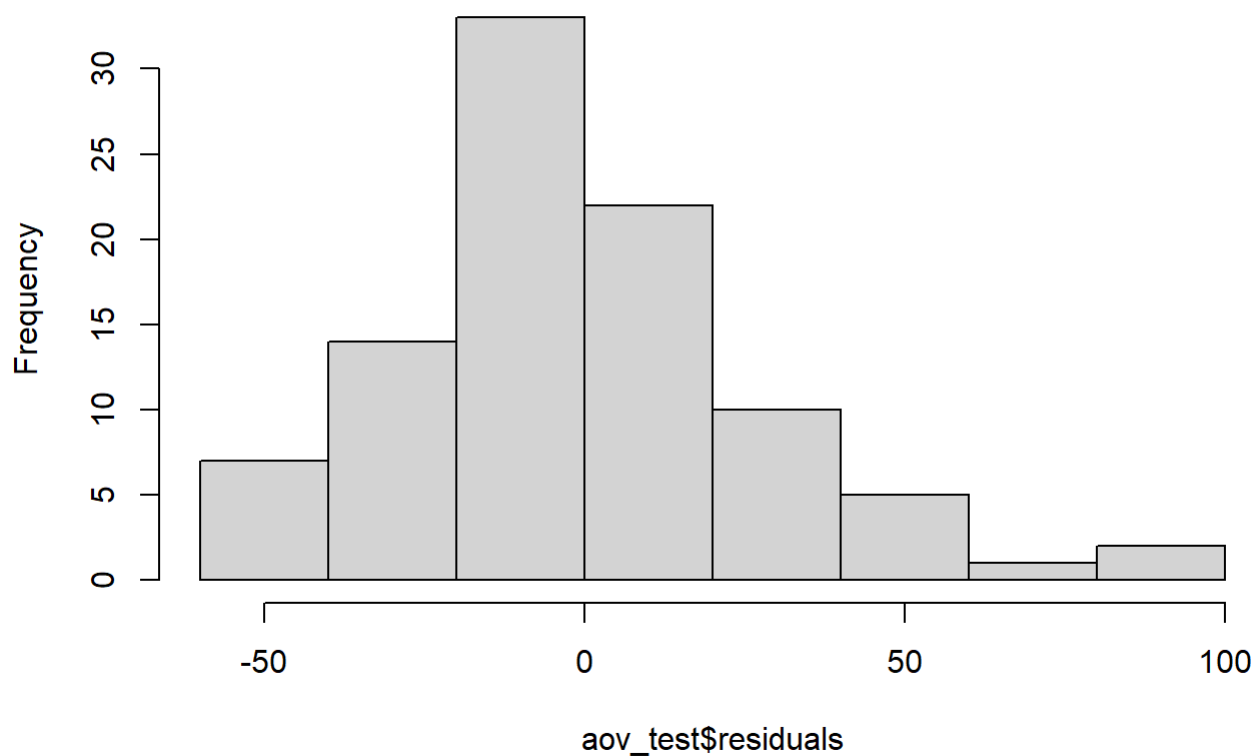
df_stack <- rbind(df1, df2)
df_stack$Southern <- as.factor(df_stack$Southern)

aov_test <- aov(Males ~ Southern, data=df_stack)
summary(aov_test)
```

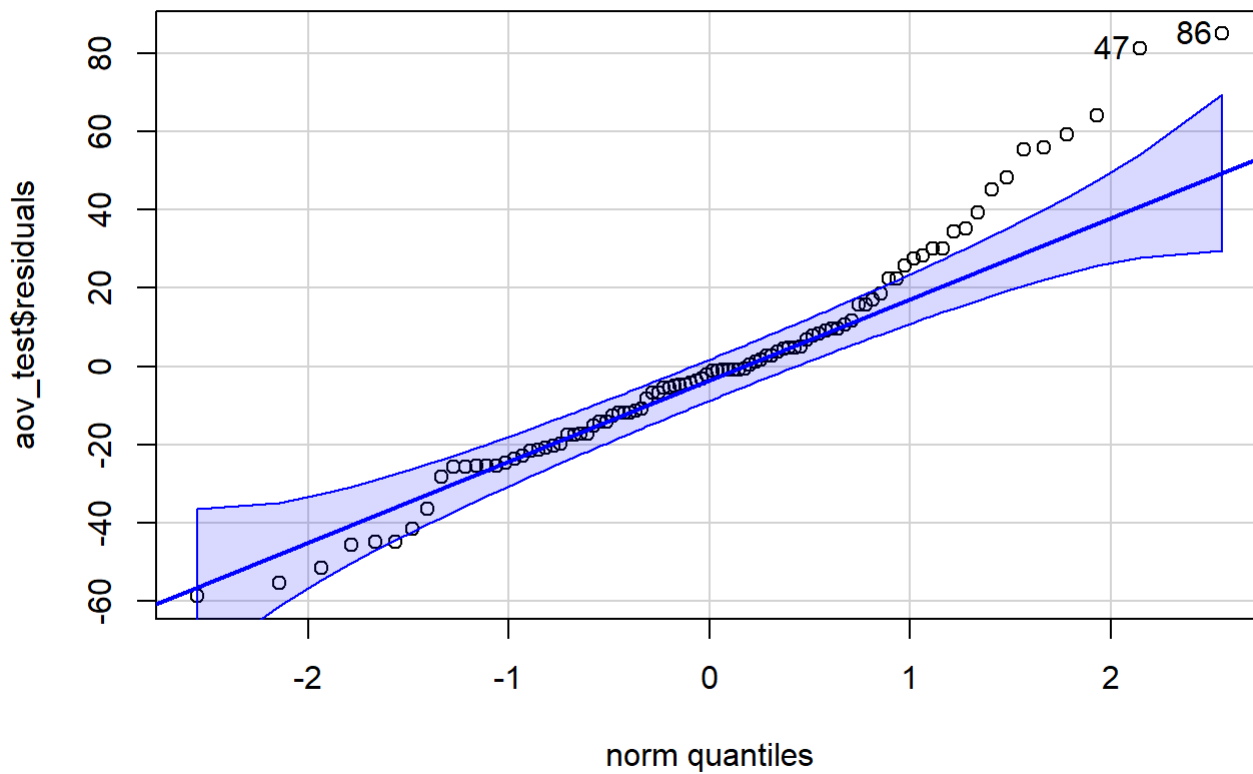
```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Southern    3   8765   2921.8    3.627  0.016 *
## Residuals   90  72501    805.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
hist(aov_test$residuals)
```

**Histogram of aov\_test\$residuals**



```
qqPlot(aov_test$residuals)
```



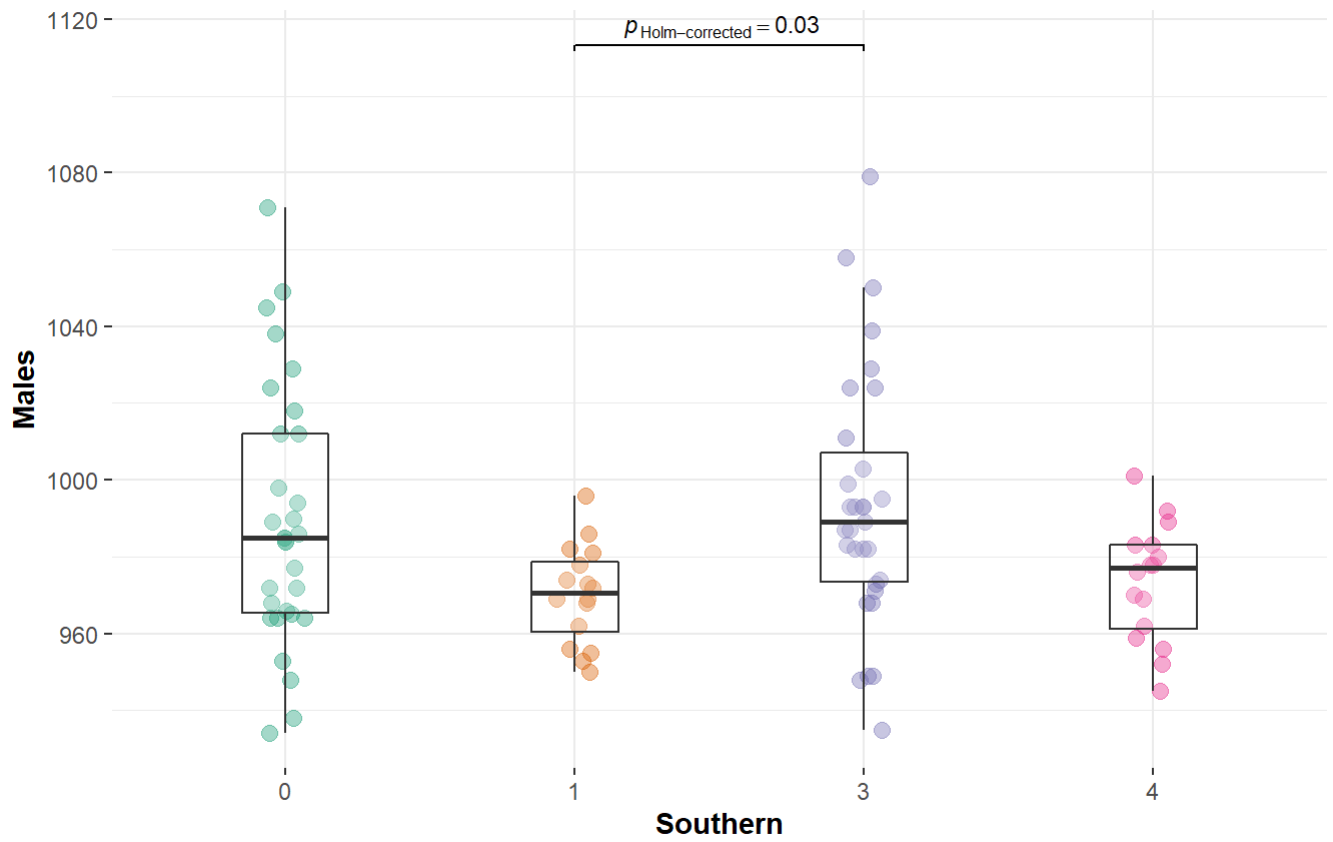
```
## [1] 86 47
```

```
kruskal.test(Males ~ Southern, data = df_stack)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  Males by Southern
## Kruskal-Wallis chi-squared = 10.624, df = 3, p-value = 0.01394
```

```
ggbetweenstats(
  data = df_stack,
  x = "Southern",
  y = "Males",
  type = "nonparametric", # ANOVA or Kruskal-Wallis
  plot.type = "box",
  pairwise.comparisons = TRUE,
  pairwise.display = "significant",
  centrality.plotting = FALSE,
  bf.message = FALSE
)
```

$\chi^2_{\text{Kruskal-Wallis}}(3) = 10.62, p = 0.01, \hat{\varepsilon}^2_{\text{ordinal}} = 0.11, \text{CI}_{95\%} [0.05, 1.00], n_{\text{obs}} = 94$



Pairwise test: **Dunn test**, Comparisons shown: **only significant**