R Notebook

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
data <- read.csv("COVID19_line_list_data.csv")

summary(data)
```

```
##        id       case_in_country   reporting.date        X
##  Min.   :   1   Min.   :   1.00   Length:1085        Mode:logical
##  1st Qu.: 272   1st Qu.:  11.00   Class :character   NA's:1085
##  Median : 543   Median :  28.00   Mode  :character
##  Mean   : 543   Mean   :  48.84
##  3rd Qu.: 814   3rd Qu.:  67.25
##  Max.   :1085   Max.   :1443.00
##                 NA's   :197
##    summary            location           country             gender
##  Length:1085        Length:1085        Length:1085        Length:1085
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##       age         symptom_onset      If_onset_approximated hosp_visit_date
##  Min.   : 0.25   Length:1085        Min.   :0.0000        Length:1085
##  1st Qu.:35.00   Class :character   1st Qu.:0.0000        Class :character
##  Median :51.00   Mode  :character   Median :0.0000        Mode  :character
##  Mean   :49.48                      Mean   :0.0429
##  3rd Qu.:64.00                      3rd Qu.:0.0000
##  Max.   :96.00                      Max.   :1.0000
##  NA's   :242                        NA's   :525
##  exposure_start     exposure_end       visiting.Wuhan    from.Wuhan
##  Length:1085        Length:1085        Min.   :0.000   Min.   :0.0000
##  Class :character   Class :character   1st Qu.:0.000   1st Qu.:0.0000
##  Mode  :character   Mode  :character   Median :0.000   Median :0.0000
##                                        Mean   :0.177   Mean   :0.1443
##                                        3rd Qu.:0.000   3rd Qu.:0.0000
##                                        Max.   :1.000   Max.   :1.0000
##                                                        NA's   :4
##    death            recovered          symptom             source
##  Length:1085        Length:1085        Length:1085        Length:1085
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
```

```
##
##
##      link              X.1             X.2             X.3             X.4
##  Length:1085        Mode:logical   Mode:logical   Mode:logical   Mode:logical
##  Class :character   NA's:1085      NA's:1085      NA's:1085      NA's:1085
##  Mode  :character
##
##
##
##
##    X.5             X.6
##  Mode:logical   Mode:logical
##  NA's:1085      NA's:1085
##
##
##
##
##
```

```
# cleaned up death column

data$death_colmn <- as.integer(data$death != 0)

# death rate
sum(data$death_colmn) / nrow(data)
```

```
## [1] 0.05806452
```

```
# AGE
# claim: people who die are older
dead = subset(data, death_colmn == 1)
alive = subset(data, death_colmn == 0)
mean(dead$age, na.rm = TRUE)
```

```
## [1] 68.58621
```

```
mean(alive$age, na.rm = TRUE)
```

```
## [1] 48.07229
```

```
# is this statistically significant?
t.test(alive$age, dead$age, alternative="two.sided", conf.level = 0.99)
```

```
##
##  Welch Two Sample t-test
##
## data:  alive$age and dead$age
## t = -10.839, df = 72.234, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  -25.52122 -15.50661
## sample estimates:
## mean of x mean of y
##  48.07229  68.58621
```

```
# normally, if p-value < 0.05, we reject null hypothesis
# here, p-value ~ 0, so we reject the null hypothesis and
# conclude that this is statistically significant
```

```r
# GENDER
# claim: gender has no effect
men = subset(data, gender == "male")
women = subset(data, gender == "female")
mean(men$death_colmn, na.rm = TRUE)
```

```
## [1] 0.08461538
```

```r
mean(women$death_colmn, na.rm = TRUE)
```

```
## [1] 0.03664921
```

```r
# is this statistically significant?
t.test(men$death_colmn, women$death_colmn, alternative="two.sided", conf.level = 0.99)
```

```
##
##  Welch Two Sample t-test
##
## data:  men$death_colmn and women$death_colmn
## t = 3.084, df = 894.06, p-value = 0.002105
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  0.007817675 0.088114665
## sample estimates:
##  mean of x  mean of y
## 0.08461538 0.03664921
```

```r
# 99% confidence: men have from 0.8% to 8.8% higher chance
# of dying.
# p-value = 0.002 < 0.05, so this is statistically
# significant
```

Add a new chunk by clicking the *Insert Chunk* button on the toolbar or by pressing *Ctrl+Alt+I*.

When you save the notebook, an HTML file containing the code and output will be saved alongside it (click the *Preview* button or press *Ctrl+Shift+K* to preview the HTML file).

The preview shows you a rendered HTML copy of the contents of the editor. Consequently, unlike *Knit*, *Preview* does not run any R code chunks. Instead, the output of the chunk when it was last run in the editor is displayed.