

# success\_failure

Owais Khan

11/17/2021

```
library(tidyr)
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.3.1 --
## v ggplot2 3.3.4      v dplyr 1.0.7
## v tibble 3.1.2       v stringr 1.4.0
## v readr 1.4.0        v forcats 0.5.1
## v purrr 0.3.4

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(reshape)

##
## Attaching package: 'reshape'

## The following object is masked from 'package:dplyr':
##
##      rename

## The following objects are masked from 'package:tidyr':
##
##      expand, smiths

dataset <- read.csv(file = "~/STA 518/R-for-data-science/data/country_success_failure.txt",
                    header = TRUE, sep = ",")
```

I can import data from a variety of sources.

I can write comments that explain the “why” (Why did you choose this approach instead of an alternative? What else did you try that didn’t work?) of my code.

I can implement resampling methods to make conclusions about data. Make the data wider by transforming column variable into row

```
#I had to transform the data into wider shape so that I can have access to each variable in the same row
wide_data <- pivot_wider(dataset, id_cols = c("ID", "NAME"), names_from = "variable", values_from = "es")
```

Rename variable ID to Id, NAME to Country, B01001\_001 to Total, B01001\_002 to Success & B01001\_026 to Failure

```
names(wide_data) <- c("Id", "Country", "Total", "Success", "Failure")
```

Calculate the success and failure rate from the respective variables.

```
wide_data <- wide_data %>% mutate(Success_Rate = (Success/Total)*100, Failure_Rate = (Failure/Total)*100)
print(wide_data)
```

```
## # A tibble: 10 x 7
##       Id Country      Total Success Failure Success_Rate Failure_Rate
##   <int> <chr>      <int>   <int>   <int>      <dbl>      <dbl>
## 1 1001020100 United States    1993     907    1086      45.5      54.5
## 2 1001020200 United Kingdom    1959    1058     901      54.0      46.0
## 3 1001020300 United Arab Emira~    3507    1731    1776      49.4      50.6
## 4 1001020400 Canada          3878    1949    1929      50.3      49.7
## 5 1001020500 Brazil         10596    5256    5340      49.6      50.4
## 6 1001020600 China           3668    1677    1991      45.7      54.3
## 7 1001020700 Pakistan         3586    1627    1959      45.4      54.6
## 8 1001020801 Russia           3196    1567    1629      49.0      51.0
## 9 1001020802 Australia        11142    5315    5827      47.7      52.3
## 10 1001020900 Turkey           6143    3123    3020      50.8      49.2
```

```
#Mean success and failure rates throughout the countries
```

```
means <- wide_data %>% summarise(mean_success = mean(Success_Rate), mean_failure = mean(Failure_Rate))
print(means)
```

```
## # A tibble: 1 x 2
##   mean_success mean_failure
##   <dbl>      <dbl>
## 1    48.7      51.3
```