

Cost of living

Willem

2022-06-30

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.6      v purrr   0.3.4
## v tibble  3.1.7      v dplyr   1.0.9
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

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

```
library(janitor)
```

```
##
## Attaching package: 'janitor'
```

```
## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test
```

```
costofliving <- read_csv("costofliving.csv")
```

```
## New names:
## * ' ' -> '...1'
```

```
## Rows: 578 Columns: 9
## -- Column specification -----
## Delimiter: ","
## chr (1): City
## dbl (7): ...1, Cost.of.Living.Index, Rent.Index, Cost.of.Living.Plus.Rent.In...
## lgl (1): Rank
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
costofliving_clean<- costofliving %>%
  clean_names
```

Tidyverse function - select - filter - mutate - rename - group_by - summarize

Demonstrating select function (*which manipulates columns*)

```
costofliving_subset<-costofliving_clean %>% #This is a comment
  select(cost_of_living_index, groceries_index)
```

```
allbutcostofliving<-costofliving_clean %>%
  select(-c(cost_of_living_index, rent_index))
```

```
costofliving_clean %>%
  select(c(-cost_of_living_index))
```

```
## # A tibble: 578 x 8
##       x1 rank city                rent_index cost_of_living_~ groceries_index
##   <dbl> <lgl> <chr>                <dbl>         <dbl>         <dbl>
## 1     1  NA Hamilton, Bermuda          96.1          124.          158.
## 2     2  NA Zurich, Switzerland          69.3          102.          136.
## 3     3  NA Basel, Switzerland           49.4           NA           137.
## 4     4  NA Zug, Switzerland            72.1          102.          133.
## 5     5  NA Lugano, Switzerland          45.0           87.0          129.
## 6     6  NA Lausanne, Switzerland         59.6           92.7          123.
## 7     7  NA Beirut, Lebanon              NA            77.0          141.
## 8     8  NA Bern, Switzerland            46.1           NA           NA
## 9     9  NA Geneva, Switzerland           75.0           NA           NA
## 10    10  NA Stavanger, Norway            35.4           72.2          102.
## # ... with 568 more rows, and 2 more variables: restaurant_price_index <dbl>,
## #   local_purchasing_power_index <dbl>
```

filter (selects rows)

```
costofliving_clean %>%
  filter(cost_of_living_index<100)
```

```
## # A tibble: 481 x 9
##       x1 rank city                cost_of_living_~ rent_index cost_of_living_~
##   <dbl> <lgl> <chr>                <dbl>         <dbl>         <dbl>
## 1    15  NA Trondheim, Norway          99.4           37.7           70.5
## 2    16  NA Tromso, Norway           99.0           37.2           70.0
## 3    17  NA Reykjavik, Iceland         97.6           46.3           73.6
## 4    20  NA Tel Aviv-Yafo, Isra~       94.5           53.2           75.2
## 5    22  NA San Francisco, CA, ~       93.9          108.          101.
## 6    23  NA Oakland, CA, United~       92.9           87.8           90.5
## 7    25  NA Santa Clara, CA, Un~       89.4           90.4           89.9
```

```
## 8 26 NA Petah Tikva, Israel 88.9 32.7 62.5
## 9 27 NA Beersheba, Israel 88.8 23.1 NA
## 10 28 NA Seattle, WA, United~ 88.5 65.8 77.9
## # ... with 471 more rows, and 3 more variables: groceries_index <dbl>,
## # restaurant_price_index <dbl>, local_purchasing_power_index <dbl>
```

```
costofliving_clean %>%
  filter(cost_of_living_index>=100) ## cities that are more or as expensive as New York
```

```
## # A tibble: 10 x 9
##       x1 rank city cost_of_living_~ rent_index cost_of_living_~
##   <dbl> <lgl> <chr>          <dbl>      <dbl>          <dbl>
## 1     2 NA Zurich, Switzerland 131.      69.3          102.
## 2     4 NA Zug, Switzerland 128.      72.1          102.
## 3     5 NA Lugano, Switzerland 124.      45.0          87.0
## 4     6 NA Lausanne, Switzerla~ 122.      59.6          92.7
## 5     7 NA Beirut, Lebanon 120.      NA           77.0
## 6     8 NA Bern, Switzerland 118.      46.1          NA
## 7    10 NA Stavanger, Norway 105.      35.4          72.2
## 8    12 NA Oslo, Norway 102.      46.4          76.1
## 9    13 NA Bergen, Norway 100.      34.8          69.7
## 10   14 NA <NA> 100      100          NA
## # ... with 3 more variables: groceries_index <dbl>,
## # restaurant_price_index <dbl>, local_purchasing_power_index <dbl>
```

```
costofliving_clean %>%
  filter(city=="Oslo, Norway") ##Will only give rows where the city is Oslo, Norway
```

```
## # A tibble: 1 x 9
##       x1 rank city cost_of_living_~ rent_index cost_of_living_~ groceries_index
##   <dbl> <lgl> <chr>          <dbl>      <dbl>          <dbl>
## 1    12 NA Oslo~ 102.      46.4          76.1          97.6
## # ... with 2 more variables: restaurant_price_index <dbl>,
## # local_purchasing_power_index <dbl>
```

```
costofliving_clean %>%
  filter(city %in% c("Oslo, Norway",
                    "Bergen, Norway",
                    "Trondheim, Norway"))
```

```
## # A tibble: 3 x 9
##       x1 rank city cost_of_living_~ rent_index cost_of_living_~ groceries_index
##   <dbl> <lgl> <chr>          <dbl>      <dbl>          <dbl>
## 1    12 NA Oslo~ 102.      46.4          76.1          97.6
## 2    13 NA Berg~ 100.      34.8          69.7          96.2
## 3    15 NA Tron~ 99.4      37.7          70.5          95.1
## # ... with 2 more variables: restaurant_price_index <dbl>,
## # local_purchasing_power_index <dbl>
```

rename

rename changes the name of a variable

```
dataset %>%  
  rename(new_variablr_name = old_variable_name)
```

```
costofliving_clean <- costofliving_clean %>%  
  rename(cofi = cost_of_living_index)
```