tutorial 2 sample code

Randall Ni

1/21/23

Introduction

This is my sample code for tutorial 2. It contains the data set of bicycle theft obtained from open data Toronto. It has been cleaned, and a graph of mean price of bike stolen in relation to hour of the day has been produced for further analysis.

Running Code

```
  ##Load libraries (prep work)
#install.packages("janitor")
library(opendatatoronto)
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   1.0.1
v tibble  3.1.7      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- 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
```

```
##Obtain data
bicycle_thefts_packages <- search_packages("bicycle thefts")

bicycle_thefts_resources <- bicycle_thefts_packages %>%
  list_package_resources()

bicycle_thefts_statistics <- bicycle_thefts_resources %>%
  filter(name == "bicycle-thefts - 4326.csv") |>
  get_resource()
```

```
##Clear Rows with NAs
bicycle_thefts_statistics <-
  na.omit(bicycle_thefts_statistics)
##Clean data, select what I wanted for graphing purposes
bicycle_thefts_statistics <-
  clean_names(bicycle_thefts_statistics)
##Select needed columns
bicycle_thefts_statistics <-
  bicycle_thefts_statistics |>
  select(report_hour, cost_of_bike)
```

```
##Wants to group by report hour
bicycle_thefts_statistics <-
  bicycle_thefts_statistics |>
  group_by(report_hour)

##Calculated the mean of the cost of bikes
bicycle_thefts_statistics <-
  bicycle_thefts_statistics |>
  summarise(
    report_hour = report_hour,
    mean_cost = mean(cost_of_bike)
    )
```

```
`summarise()` has grouped output by 'report_hour'. You can override using the
`.groups` argument.
```

```r
##cleaned out duplicated data
bicycle_thefts_statistics <-
  unique(bicycle_thefts_statistics)
```

```r
##Adding bar graph for future analysis
bicycle_thefts_statistics|>
  ggplot(aes(x = report_hour, y = mean_cost)) + geom_bar(stat = "identity", fill= "pink")
```