Stars Dataset

Caitlyn Nguyen

2023-01-31

Here we imported the dataset and broke down each star type into its own dataframe.

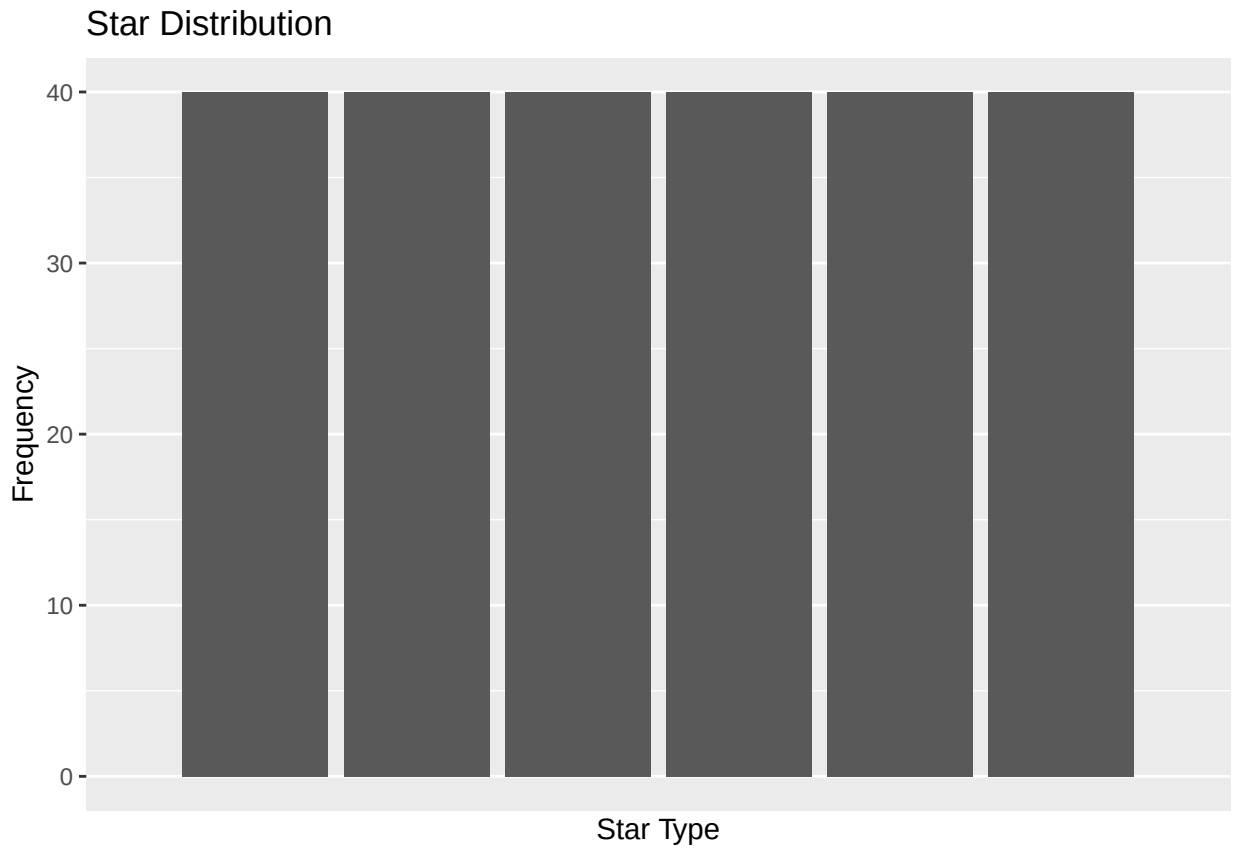
```
stars <- read.csv("stars.csv")

star_type0 <- stars[stars$Star.type == 0,]
star_type1 <- stars[stars$Star.type == 1,]
star_type2 <- stars[stars$Star.type == 2,]
star_type3 <- stars[stars$Star.type == 3,]
star_type4 <- stars[stars$Star.type == 4,]
star_type5 <- stars[stars$Star.type == 5,]
```

As we created a bar plot of the distribution of the frequency of all the different star types, we discovered it is evenly distributed. This thus indicates that the data is evenly distributed for each factor.

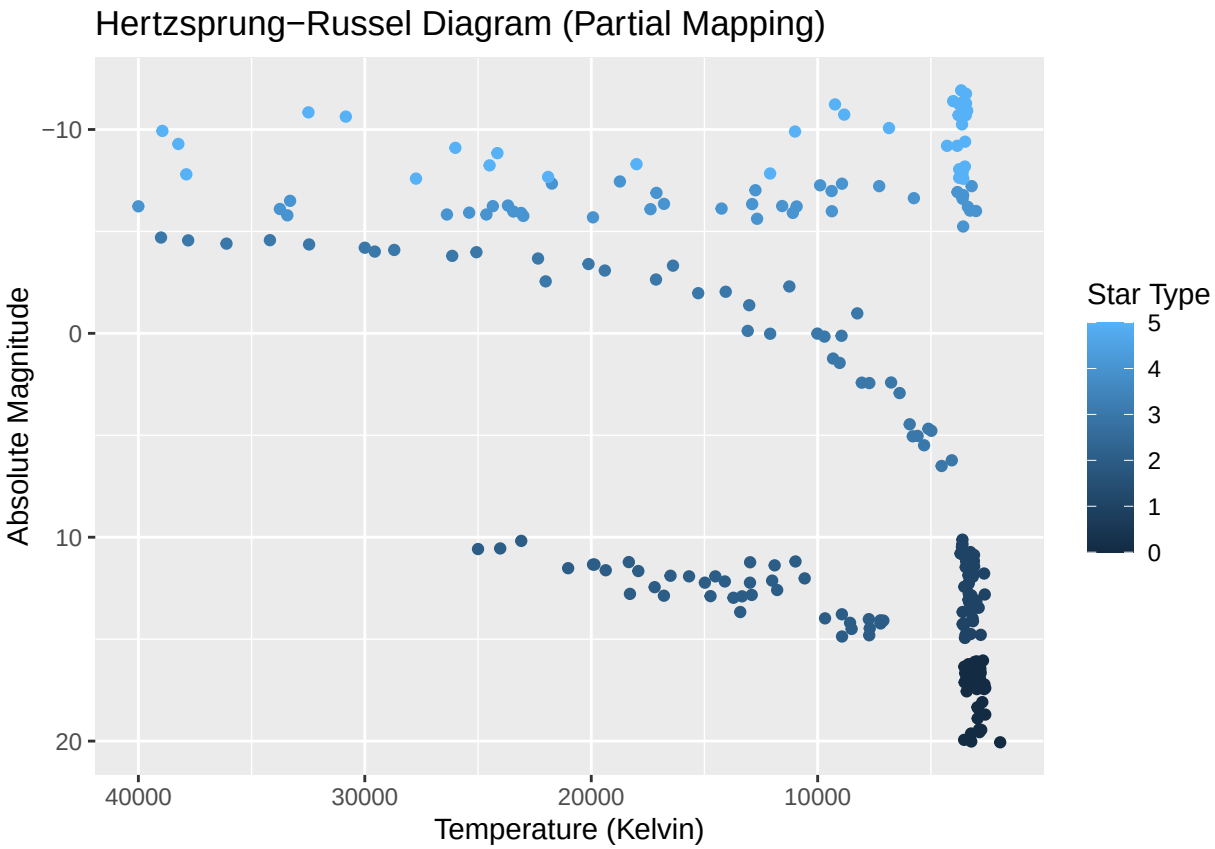
```
library(ggplot2)

ggplot(data = stars, mapping = aes(x = Star.type, fill = Star.type)) +
  geom_bar() +
  labs(x = "Star Type", y = "Frequency", title = "Star Distribution") +
    scale_x_discrete(labels = c("Brown Dwarf", "Red Dwarf", "White Dwarf",
                                "Main Sequence", "Supergiant", "Hypergiant"))
```



Here is a partial mapping of the the diagram shown in the Kaggle description.

```
ggplot(data = stars) +
  geom_point(mapping = aes(x = Temperature..K., y = Absolute.magnitude.Mv.,
                           color = Star.type)) +
  scale_x_reverse() +
  scale_y_reverse() +
  labs(x = "Temperature (Kelvin)",
       y = "Absolute Magnitude",
       title = "Hertzsprung-Russel Diagram (Partial Mapping)",
       color = c("Star Type"))
```


Hertzsprung−Russel Diagram (Partial Mapping)
Absolute Magnitude
−10
0
10
20
40000
30000
20000
10000
Temperature (Kelvin)
Star Type
5
4
3
2
1
0