Lab 02: Data wrangling

Layal Ghryani

Oct 11th 2023

Packages

```
library(tidyverse)
```

Data

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

Exercise 1

```
lego <- lego %>% filter(pieces > 0) %>% filter(!is.na(year)) %>%
  filter(!is.na(retail_price))  %>% filter (retail_price != 0)
# %>% Pipe it to the one before
```

way 1: lego_filtered <- lego %>% filter(!is.na(pieces)) %>% filter(pieces != 0) %>% filter(!is.na(retail_price)) %>% filter(retail_price != 0) %>% filter(!is.na(year))

way 2 : lego <- filter(lego, pieces > 0) lego <- filter(lego,!is.na(year)) lego <- filter(lego,!is.na(retail_price)) lego <- filter(lego,retail_price != 0) testing <- filter(lego,retail_price == 0)

Exercise 2

```
lego2 <- arrange(lego,desc(retail_price)) %>% slice(1:3) #desc = descending order
```

Another way: lego_filtered % > % arrange(desc(retail_price)) % > % slice(1:3)%>% print(width = Inf)

Describe the three most expensive sets here. the most expensive 3 lego sets are Millennium Falcon which costs 800$ and has 7541 pieces, the second most expensive set is Connections Kit which costs 755$ and has 2455 pices, the third most expensive set is Death Star whcih costs 500$ and contains 4016 pieces.

Exercise 3

```
lego <- mutate(lego,price_per_piece = retail_price/pieces)
```

Exercise 4

```
lego4<-lego %>% arrange(desc(price_per_piece)) %>% slice(1:5) %>%
  select(name,themegroup,theme,pieces,price_per_piece)
```

another way: lego %>% arrange(desc(price_per_piece)) %>% slice(1:5) %>% select(name, themegroup, theme, pieces, price_per_piece)

Describe what you notice about the sets with the highest price per piece. The highest prices are the sets consisting of 1 piece only.

Exercise 5

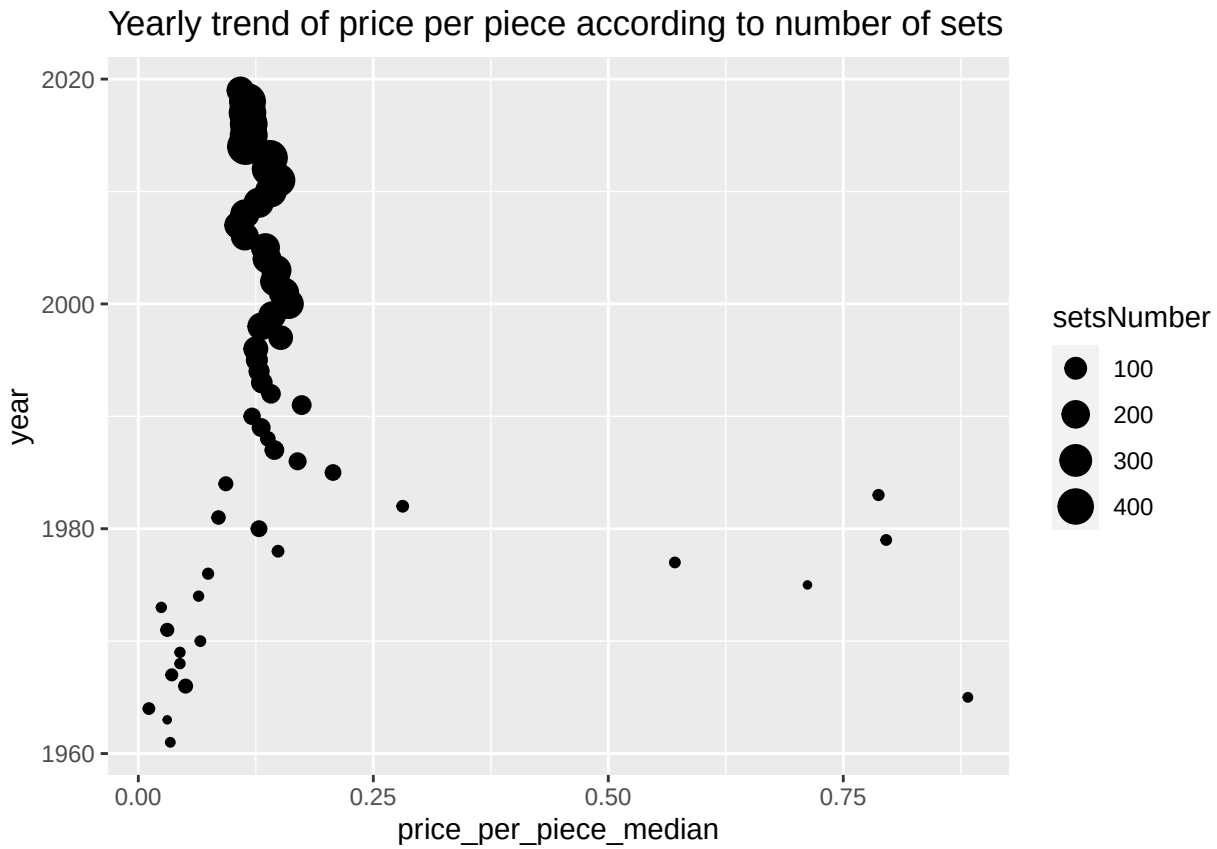
```
lego5 <- lego %>% filter(theme == "The Lord of the Rings") %>%
  group_by(subtheme) %>% summarise(maximum=max(retail_price),minimum=min(retail_price),
                                   count = n())
```

Exercise 6

```
lego6 <- group_by(lego, year)
yearly_trends <- summarize(lego6, setsNumber = n(),
                      price_per_piece_median = median(price_per_piece))
```

Exercise 7

```
library(ggplot2)
ggplot(data=yearly_trends, mapping=aes(x=price_per_piece_median,
                                       y=year, size = setsNumber)) +
  geom_point() +
  labs(title = "Yearly trend of price per piece according to number of sets")
```



Comment on what you observe in the plot above. The size of the sets increasing over the years.