Lab 3

Lucy Whitmore

2023-02-06

Getting Started

You can download the transit_cost.csv data from the website.

```
require(tidyverse)
require(lubridate)
require(ungeviz)
require(ggtext)
require(ggrepel)
require(ggforce)
library(gghighlight)

transit_cost <- read_csv('./transit_cost.csv')</pre>
```

Question 1

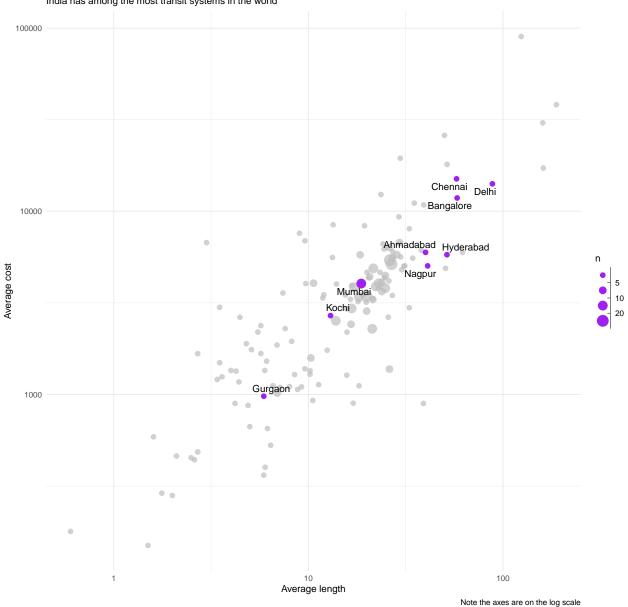
Suppose that you want to demonstrate the relationship between Average Length and Average Cost for the transit systems across all countries in the dataset. Reproduce the plot on the next page by following the procedures:

- 1. Compute the average length and average cost of transit systems by country and city
- 2. Create a basic scatter plot by placing **Average Length** on the x-axis and **Average Cost** on the y-axis.
- 3. In the scatter plot,make the size of the data points represent the number of transit systems in that particular city (Hint: use aes(size=) within the geom_point() function).
- 4. Customize the legend so it shows 5, 10, and 20 as break points for the size of data points (hint: add the feature to the plot by using scale_size_binned())
- 5. Make sure all data points are grayish except the cities from India. Make the color for the data points from these 9 cities different than the rest.
- 6. Adjust the scale of the x-axis and y-axis using the scale_y_log10() and scale_x_log10() functions so they are on the logarithmic scale.
- 7. Add the names of the cities in India using the geom_text_repel() function.
- 8. Adjust the theme settings.

```
# compute average length and average cost
transit_cost_means <- transit_cost %>%
  group_by(country, city) %>%
  mutate(real_cost = as.numeric(real_cost)) %>%
  summarise(m_length = mean(length), m_cost=mean(real_cost), n = n()) %>%
  ungroup() %>%
  na.omit()
```

```
#scatter plot
transit_cost_means %>%
    ggplot(aes(m_length, m_cost)) +
    geom_point(aes(size=n), col = "purple") +
    scale_size_binned(breaks = c(5, 10, 20)) +
    scale_x_log10() +
    scale_y_log10() +
    gghighlight(country=="IN", unhighlighted_colour = NULL) +
    geom_text_repel(data=subset(transit_cost_means, country=="IN"), aes(label=city)) +
    labs(x="Average length", y= "Average cost", caption="Note the axes are on the log scale", title ="Longthene_minimal()
```

Longer transit systems tend to cost more India has among the most transit systems in the world



Question 2

Using basically the same data, reproduce the following plot on the next page.

- 1. Compute the average length and average cost of transit systems by country and city.
- 2. Create a basic scatter plot by placing Average Length on the x-axis and Average Cost on the y-axis.
- 3. In the scatter plot,make the size of the data points represent the number of transit systems in that particular city (Hint: use aes(size=) within the geom_point() function).
- 4. Customize the legend so it shows 5, 10, and 20 as break points for the size of data points (hint: add the feature to the plot by using scale_size_binned())
- 5. Make sure all data points are grayish except the cities from US. Make the color for the data points from the US cities different than the rest.
- 6. Adjust the scale of the x-axis and y-axis using the scale_y_log10() and scale_x_log10() functions so they are on the logarithmic scale.
- 7. Using the geom_mark_ellipse() function from the ggforce package, circle the data points for the US cities.
- 8. Add the names of the US cities using the geom_label_repel() function.
- 9. Adjust the theme settings.

```
#scatter plot
transit_cost_means %>%
    ggplot(aes(m_length, m_cost)) +
    geom_point(aes(size=n), color="lightblue") +
    scale_size_binned(breaks = c(5, 10, 20)) +
    scale_x_log10() +
    scale_y_log10() +
    gghighlight(country="US", unhighlighted_colour = NULL) +
    geom_label_repel(data=subset(transit_cost_means, country=="US"), aes(label=city)) +
    geom_mark_ellipse(aes(m_length, m_cost, filter = country =="US", color="red")) +
    labs(x="Average length", y= "Average cost", caption="Note the axes are on the log scale", title ="Longtheme_minimal()
```

Longer transit systems tend to cost more



