

A33 Exercise

Daniella Opoku Darkwah

1/28/2024

Task 1: Reflection

Engaging with Amanda Cox's keynote was an eye-opener, shedding light on the nuanced challenge of communicating uncertainty in data, a crucial aspect of my role as an upcoming business analyst and data visualization expert. Nicole Torres' exploration of visualizing uncertainty and the discussion on histograms and density plots further deepened my understanding. Cox's emphasis on transparency in scenarios with life-altering implications struck a chord, making me contemplate the gravity of our responsibility.

Recently, I encountered a data visualization that exemplified seamless clarity, while another fell prey to unnecessary complexity, hindering rather than aiding comprehension. Applying these insights to my own experiences, I have learned to balance aesthetics with transparency. Striving for simplicity without sacrificing informativeness is now a focal point. Going forward, I aim to integrate these principles into my visualizations, ensuring stakeholders not only find the visuals engaging but also leave with a clear grasp of the uncertainties inherent in the data.

Task 2: Visualizing uncertainty with gapminder

For this exercise you'll revisit Hans Rosling's gapminder data on health and wealth. Install the **gapminder** package if you haven't already. Also install the **ggribes** and **ggshaves** packages.

```
library(tidyverse) # For ggplot, dplyr, and friends
library(gapminder) # For health/wealth data
library(ggribes)    # For ridge plots
library(ggshaves)   # For half geoms
```

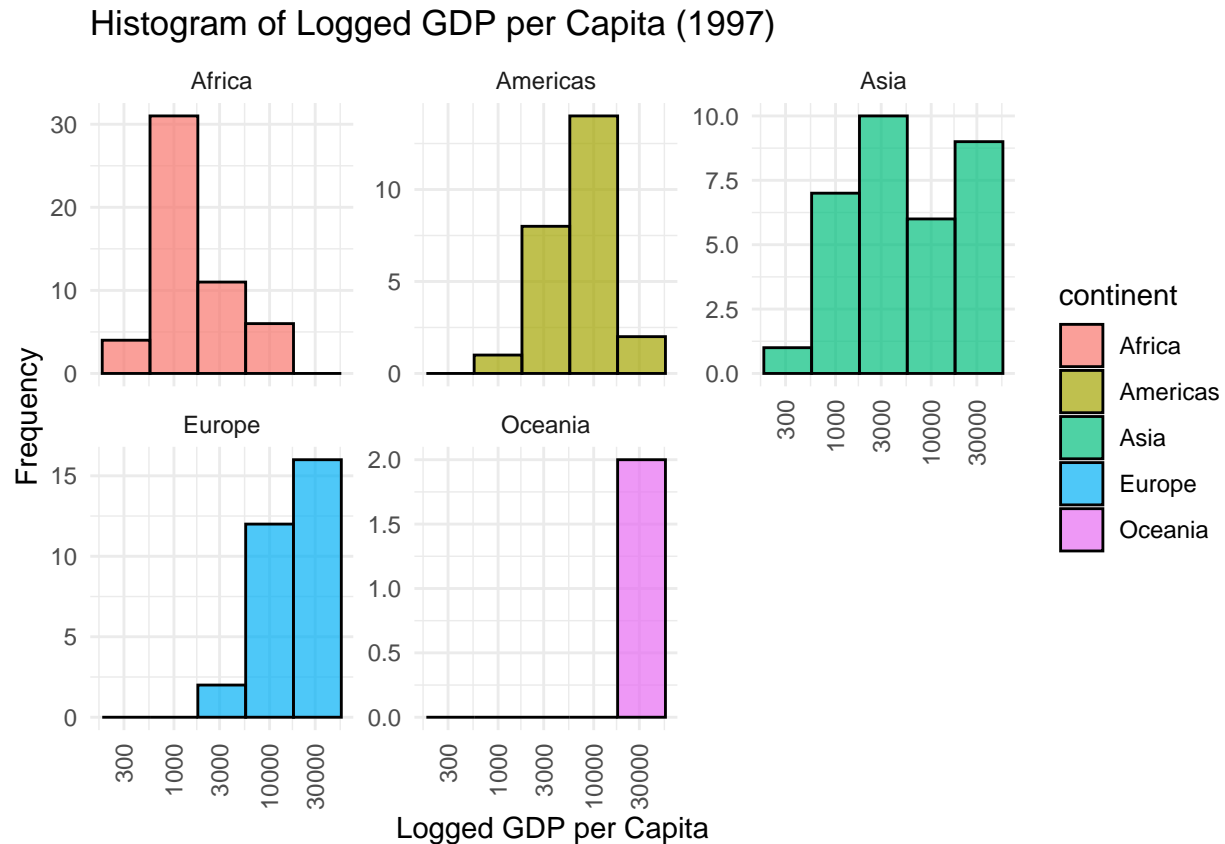
Histograms

Make a histogram of *logged* GDP per capita for 1997 only, across all five continents.

```
# Create a filtered dataset
gapminder_1997 <- gapminder %>%
  filter (year == 1997)

# Add a histogram here. You'll probably want to fill and facet.
# You'll also need to add scale_x_log10()
ggplot(gapminder_1997, aes(x = gdpPercap, fill = continent)) +
  geom_histogram(binwidth = 0.5, color = "black", alpha = 0.7, position = "identity") +
```

```
scale_x_log10() +
labs(title = "Histogram of Logged GDP per Capita (1997)",
x = "Logged GDP per Capita",
y = "Frequency") +
facet_wrap(~continent, scales = "free_y") +
theme_minimal() +
theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1))
```



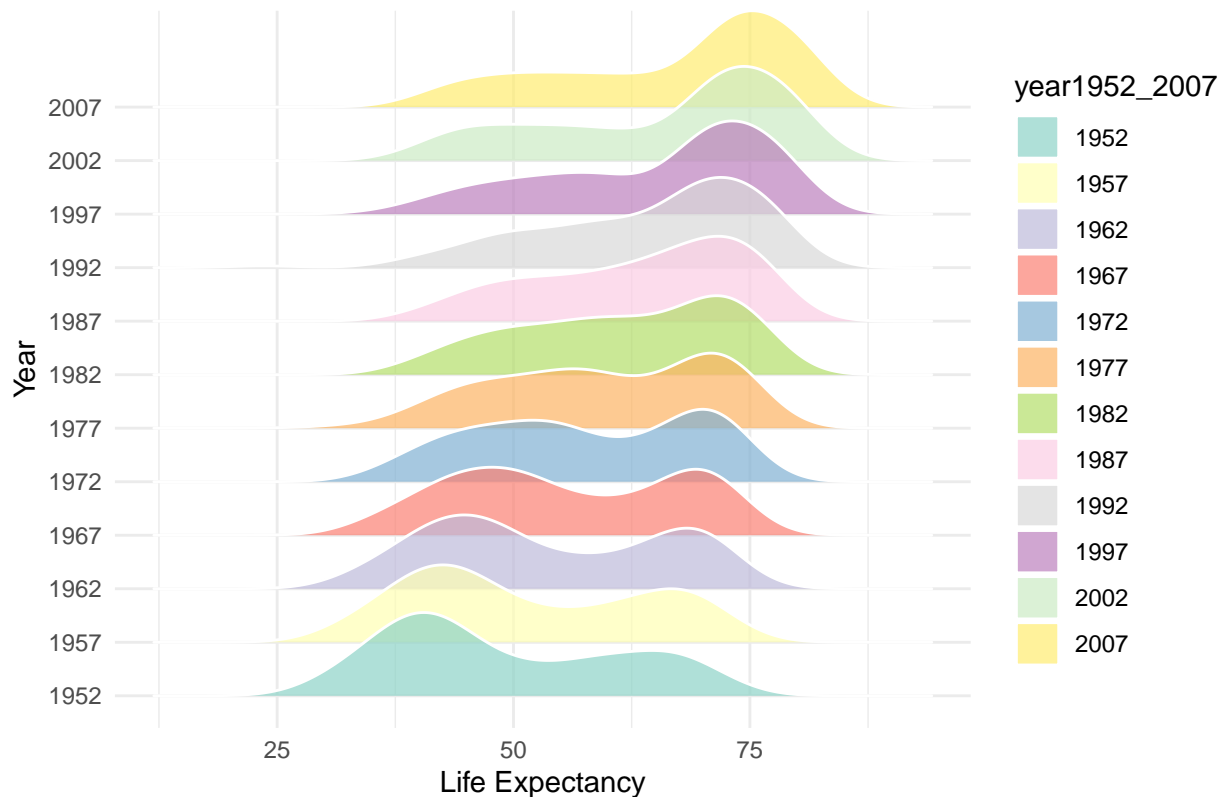
Densities

Make a ridge plot of life expectancy over time, from 1952 to 2007. You'll need to use the full gapminder data, not the 1997-only data.

```
# Add a ridge plot here
# Create a categorical year variable
gapminder <- gapminder %>%
mutate(year1952_2007 = as.factor(year))
# Create a ridge plot
ggplot(gapminder, aes(x = lifeExp, y = year1952_2007, fill = year1952_2007)) +
geom_density_ridges(alpha = 0.7, color = "white") +
scale_fill_brewer(palette = "Set3") +
labs(title = "Ridge Plot of Global Life Expectancy Over Time (1952-2007)",
x = "Life Expectancy",
y = "Year") +
theme_minimal()
```

```
## Picking joint bandwidth of 3.88
```

Ridge Plot of Global Life Expectancy Over Time (1952–2007)



Boxes, violins, and others

Make a filtered dataset that selects data from only 2007 *and* removes Oceania. Show the distribution of *logged* GDP per capita across the *four* continents using some combination of boxplots and/or violin plots and/or strip plots, either overlaid on top of each other, or using their `geom_half_*()` counterparts from `gghalves`.

```
# Create a filtered dataset
gapminder_2007 <- gapminder %>%
  filter (year == 2007, continent != "Oceania")

# Add a plot here. You'll want to use either scale_x_log10() or scale_y_log10(),
# depending on which direction your boxplots/violin plots/etc.
ggplot(gapminder_2007, aes(x = continent, y = log(gdpPercap))) +
  geom_violin(trim = FALSE, fill = "#99CCFF", color = "blue") +
  geom_boxplot(width = 0.2, fill = "#FF6666", color = "red", outlier.shape = NA) +
  geom_jitter(position = position_jitter(width = 0.15), color = "black", size = 0.5) +
  labs(x = "Continent", y = "Logged GDP per Capita",
  title = "Distribution of Logged GDP per Capita (2007) Across Continents") +
  theme_minimal() +
  scale_y_log10()
```

Distribution of Logged GDP per Capita (2007) Across Continents

