

# Project 5

Read in the dataset you will be working with:

```
outer_space_objects <- read.csv("outer_space_objects.csv")
paged_table(outer_space_objects)
```

Entity	Code	Year	num_objects
<chr>	<chr>	<int>	<int>
APSCO		2023	1
Algeria	DZA	2002	1
Algeria	DZA	2010	1
Algeria	DZA	2016	3
Algeria	DZA	2017	1
Angola	AGO	2017	1
Angola	AGO	2022	1
Arabsat		1985	2
Arabsat		1992	1
Arabsat		1996	2

1-10 of 1,175 rows

Previous123456...118Next

More information about the dataset can be found here: <https://github.com/rfordatascience/tidytuesday/blob/main/data/2024/2024-04-23/readme.md> and here: <https://www.unoosa.org/oosa/osoindex/search-ng.jspx>

**Question:** Who leads the space race?

**Introduction:** The dataset used for this analysis contains information about the annual number of objects launched into outer space. Objects are defined as satellites, probes, landers, crewed spacecrafts, and space station flight element launched into Earth orbit or beyond. Each row corresponds to the annual number of objects a country or an organization has launched in a given year. If a country did not conduct any launches in a specific year, it is excluded from the dataset for that year.

The dataset requires some cleaning before creating visuals and gaining insights. The question that will be answered from this analysis is the following: “Who leads the space race?”. To answer the question, three key variables will be needed. The variables are the following: `Entity`, `Year`, `num_objects`. `Entity` is a character variable indicating the name of the country or the organization which launched the objects. `Year` is an integer variable and represents the exact year the objects were launched into space. Another integer variable, `num_objects`, indicates the annual number of objects launched into outer space.

**Approach:** To address the question, countries will be grouped into seven distinct categories based on the number of space launches over the last 20 years. These categories will be visualized on a world map, making it easy for readers to observe the variation in space activity across different countries and continents. The map will use carefully chosen colors to emphasize the differences between categories, aiding comprehension.

Additionally, a line graph will highlight trends in space launches over the past two decades for the leading countries and unions (such as Europe) that account for the largest share of launches. Finally, a summary table will provide a concise overview of the data, allowing for quick comparison and reinforcing the insights from the visualizations.

**Analysis:**

```

# Load the world map as an sf object
world_sf <- ne_countries(scale = "medium", returnclass = "sf") %>%
  filter(name != "Antarctica") # Remove Antarctica

# Load the dataset
outer_space_objects <- read.csv("outer_space_objects.csv")

# Filter and summarize data for the years 2004 to 2023
outer_space_objects_summarized <- outer_space_objects %>%
  filter(Year >= 2004 & Year <= 2023) %>% # Filter for the desired years
  group_by(Entity) %>% # Group by entity (country/organization)
  summarise(Total_Objects = sum(num_objects, na.rm = TRUE)) # Sum objects for each entity

# Create a mapping table for unmatched names
mapping <- data.frame(
  Entity = c("United States", "United Kingdom", "Czechia"),
  name = c("United States of America", "United Kingdom", "Czech Republic")
)

# Join the mapping table to align entity names
outer_space_objects_summarized <- outer_space_objects_summarized %>%
  left_join(mapping, by = "Entity") %>%
  mutate(
    name = ifelse(is.na(name), Entity, name) # Use `Entity` if no mapping exists
  )

# Categorize the data and define the correct order of the categories
outer_space_objects_summarized <- outer_space_objects_summarized %>%
  mutate(
    Category = factor(
      case_when(
        Total_Objects <= 5 ~ "1-5",
        Total_Objects <= 10 ~ "6-10",
        Total_Objects <= 50 ~ "11-50",
        Total_Objects <= 200 ~ "51-200",
        Total_Objects <= 500 ~ "201-500",
        Total_Objects > 500 ~ "Over 500"
      ),
      levels = c("1-5", "6-10", "11-50", "51-200", "201-500", "Over 500", "NA") # Define correct order
    )
  )

# Join the summarized data with the world map
map_data <- world_sf %>%
  left_join(outer_space_objects_summarized, by = c("name"))

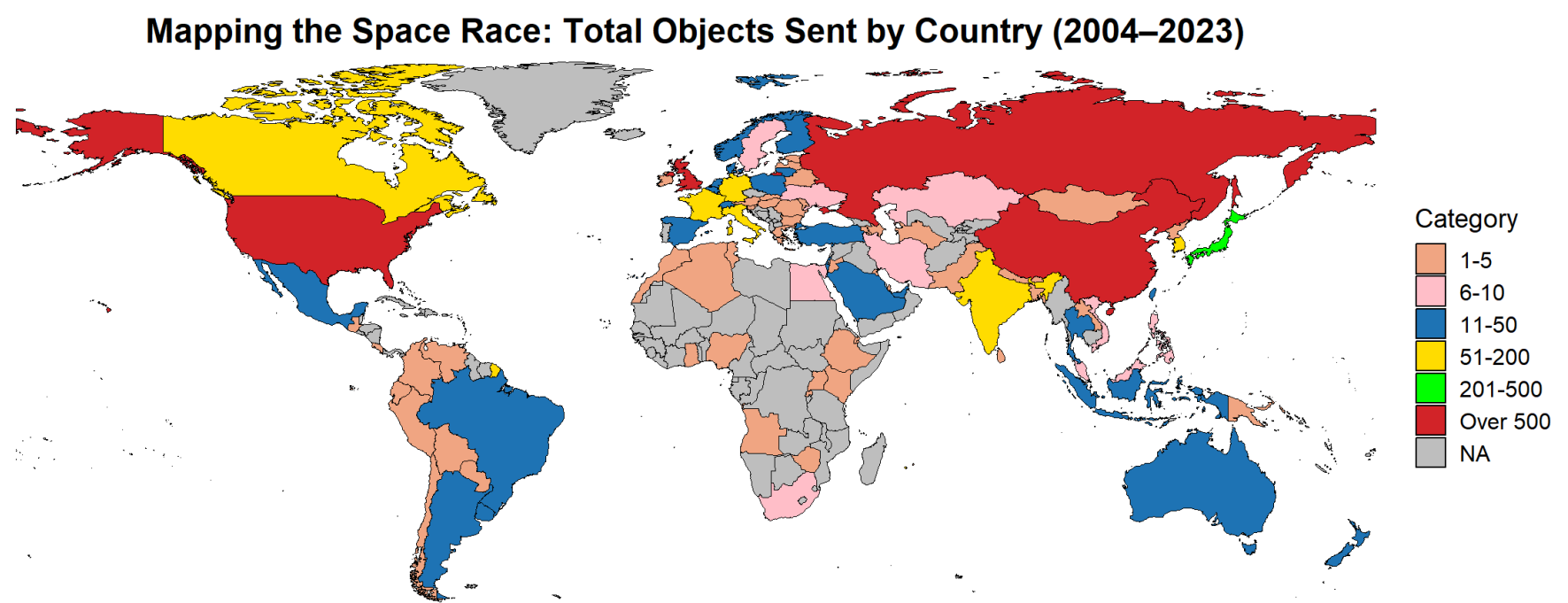
# Create the map with distinct categories
ggplot(map_data) +
  geom_sf(aes(fill = Category), color = "black", size = 0.1) +
  scale_fill_manual(
    values = c(
      "1-5" = "#f4a582",
      "6-10" = "pink",
      "11-50" = "#1f77b4",
      "51-200" = "#ffdd00",
      "201-500" = "#00ff00",
      "Over 500" = "#d62728",
      "NA" = "gray"
    ),
    na.value = "gray"
  ) +
  coord_sf(expand = FALSE) +
  theme_minimal(base_size = 14) +
  labs(
    title = "Mapping the Space Race: Total Objects Sent by Country (2004-2023)",
  )

```

```

    fill = "Category",
    x = "",
    y = ""
  ) +
  theme(
    plot.title = element_text(hjust = 0.5, size = 18, face = "bold"),
    legend.text = element_text(size = 12),
    legend.title = element_text(size = 14),
    axis.text = element_blank(),
    axis.ticks = element_blank(),
    panel.grid = element_blank()
  )

```



```

# Define the List of European countries excluding Russia and United Kingdom
european_countries <- c(
  "France", "Germany", "Italy", "Spain", "Poland", "Sweden", "Netherlands",
  "Belgium", "Austria", "Ireland", "Finland", "Denmark", "Czechia", "Portugal",
  "Greece", "Hungary", "Slovakia", "Slovenia", "Bulgaria", "Croatia", "Estonia",
  "Latvia", "Lithuania", "Luxembourg", "Malta", "Cyprus", "Romania"
)

# Add a group column for Europe and other countries
filtered_data <- outer_space_objects %>%
  mutate(
    Group = case_when(
      Entity == "United States" ~ "United States",
      Entity == "United Kingdom" ~ "United Kingdom",
      Entity == "Russia" ~ "Russia",
      Entity == "China" ~ "China",
      Entity %in% european_countries ~ "Europe",
      TRUE ~ NA_character_ # Exclude others
    )
  ) %>%
  filter(Year >= 2004 & Year <= 2023, !is.na(Group)) # Remove invalid groups

# Summarize the number of objects by year and group
grouped_data <- filtered_data %>%
  group_by(Year, Group) %>%
  summarise(num_objects = sum(num_objects, na.rm = TRUE), .groups = "drop")

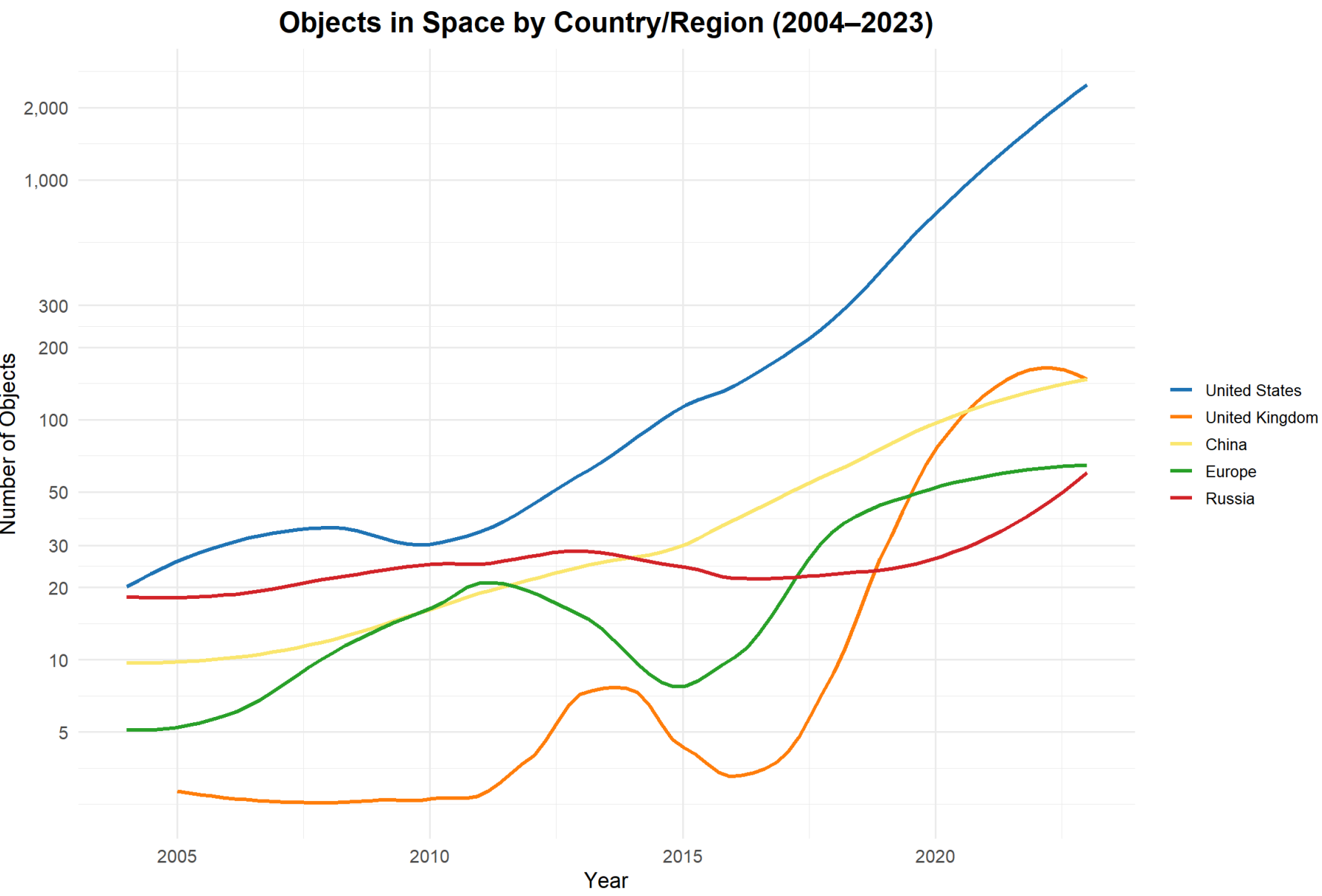
grouped_data <- grouped_data %>%
  mutate(Group = factor(Group, levels = c("United States", "United Kingdom", "China", "Europe", "Russia")))

# Plot with reordered Legend
ggplot(grouped_data, aes(x = Year, y = num_objects, color = Group, group = Group)) +
  geom_smooth(se = FALSE, method = "loess", span = 0.5, size = 1.2) + # Smooth Lines
  scale_y_continuous(
    trans = 'log10',
    breaks = c(0, 5, 10, 20, 30, 50, 100, 200, 300, 1000, 2000), # Custom breaks
    labels = scales::comma # Format Labels
  ) +
  # Manual color scale for Lines
  scale_color_manual(
    values = c(
      "United States" = "#1f77b4",
      "United Kingdom" = "#ff7f0e",
      "China" = "#f7b6d2",
      "Europe" = "#2ca02c",
      "Russia" = "#d62728"
    )
  ) +
  theme_minimal(base_size = 14) +
  labs(
    title = "Objects in Space by Country/Region (2004-2023)",
    x = "Year",
    y = "Number of Objects",
    color = "Group" # Set Legend title for color
  ) +
  theme(
    plot.title = element_text(hjust = 0.5, size = 18, face = "bold"),
    legend.position = "right", # Move the Legend to the right
    legend.title = element_blank(),
    legend.text = element_text(size = 10)
  )

```

```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



```
specific_years <- c(2004, 2009, 2014, 2019, 2023)

# Summarize data for specific years
summary_table <- filtered_data %>%
  filter(Year %in% specific_years) %>%
  group_by(Group, Year) %>%
  summarise(num_objects = sum(num_objects, na.rm = TRUE), .groups = "drop") %>%
  pivot_wider(names_from = Year, values_from = num_objects, values_fill = 0)

# View the table
print(summary_table)
```

```
## # A tibble: 5 × 6
##   Group      `2004` `2009` `2014` `2019` `2023`
##   <chr>      <int>  <int>  <int>  <int>  <int>
## 1 China         11     8    25    82   128
## 2 Europe         9    10    15    40    59
## 3 Russia        19    30    35    30    62
## 4 United Kingdom  0     3    10    14   144
## 5 United States  21    38    95   362  2166
```

**Discussion:** When analyzing the visualization and accompanying data, several trends emerge regarding global space launch activity over the past 20 years. Asia, Europe, and North America lead with the highest number of launches, while South America, Africa, and Oceania show significantly lower activity. Africa records the fewest launches, with only a handful across the entire continent. Notably, only four countries exceed the 500-launch mark, highlighting the concentration of space activity among a small number of nations. Many countries have yet to conduct any space launches.

The line graph with the biggest contributors over the last 20 years provides some helpful insights. The United Kingdom had almost no launches between 2005 and 2013, but after 2017, a steep increase in yearly launches placed the country as the second biggest contributor to space launches. China has shown a linear increase over the last 20 years, and from 2014, it has been in the top 3 contributors. Russia was the second biggest contributor between 2005 and 2014, but after that, China, Europe, and the United Kingdom surpassed Russia and ranked it as the 5th biggest contributor in 2023. Europe has shown fluctuations over the years, follows a similar trend to the United Kingdom, and experienced a steep increase in 2017. In 2023, it ranked as the 4th biggest contributor. Last but not least, the United States has led space launches over the last 20 years. Since 2017, it has shown exponential growth, and in 2023, it recorded 15 times more contributions than the second biggest contributor.

The summary table helps in making comparisons between the 5 biggest contributors. In 2004 and 2009, the United States was barely surpassing Russia with a few launches, while China, Europe, and the United Kingdom were far behind. By 2014, the United States had more launches than all the other four contributors combined. This trend continued in the following years, with 2019 and 2023 recording 2 and 5 times more launches than all the other contributors combined, respectively.

In conclusion, the United States is leading the space race, while Europe and the United Kingdom have shown a significant increase in their space launches over the last decade. China has exhibited a linear increase in space launches over the last two decades, and Russia remains consistently one of the biggest contributors. All contributors must significantly increase their space launches to keep pace with the United States' exponential growth.