

# SPEC REU R Resources: Basic Data Visualization with ggplot2 – Homework

Marie Zaragoza, Claudia Salas Gimenez and Ben Graham

Summer 2024

Welcome to your final assignment in the Data Visualization I module. By now, you’ve been introduced to the fundamentals of data visualization in R using `ggplot2` and learned how to custom aesthetics to create publication-quality visuals. This assignment will bring together those skills, encouraging you to create polished visualizations—just as you would in your own research—by practicing with histograms and kernel density plots, and for those looking for a challenge, designing your own visualizations.

You’ll be using replication data from the following paper: Markowitz, Jonathan, Suzie Caldwell Mulesky, Benjamin A.T. Graham, and Christopher J. Fariss. 2020. “Productive Pacifists: The Rise of Production-Oriented States and Decline of Profit-Motivated Conquest.” *International Studies Quarterly*, 64(3): 558-572.

Save your responses in your personal subfolder in the 412\_413 shared AY24-25 Google Drive folder. The R script should be titled `HW_DV1_[YOUR INITIALS]`. You can also save a copy of your R script to your own computer for future reference.

## Initial Setup

First, set your R environment to the location of the data files, and load the `tidyverse` and `ggplot2` packages and the necessary data. In this homework assignment, you’ll be using replication data `ProductivePacifists_Data.RDATA` from the paper:

Markowitz, Jonathan, Suzie Caldwell Mulesky, Benjamin A.T. Graham, and Christopher J. Fariss. 2020. “Productive Pacifists: The Rise of Production-Oriented States and Decline of Profit-Motivated Conquest.” *International Studies Quarterly*, 64(3): 558-572.

```
# Set working directory
#setwd("YourFolderPath")

# Load required libraries
library(ggplot2)
library(tidyverse)

# Load the dataset
load("ProductivePacifists_Data.RDATA")
```

Attached is the full research paper “[Productive Pacifists: The Rise of Production-Oriented States and Decline of Profit-Motivated Conquest](#)” and the [codebook](#) for better understand the structure and variables within the data. Now, let’s get started!

## Exercise 1: GDP Per Capita Histogram

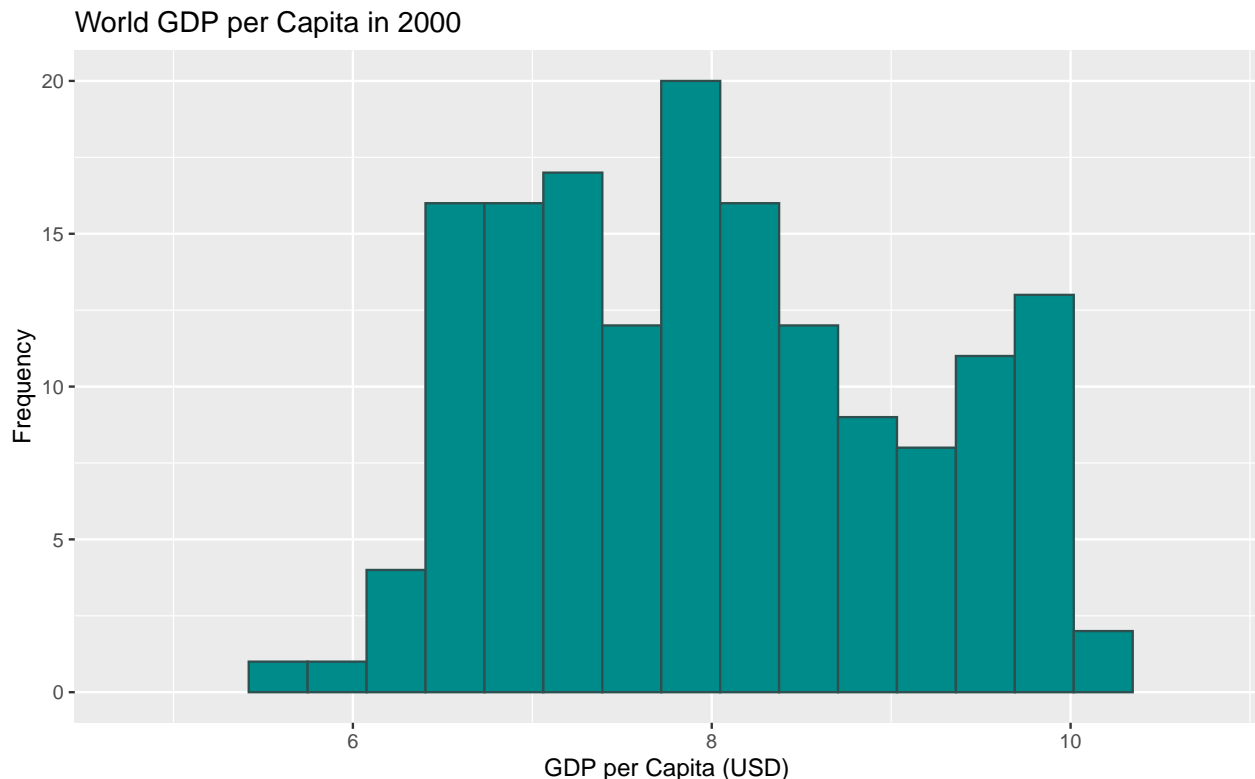
Start by generating a histogram of GDP per capita for the year 2000. At a minimum, your plot should be colored and include an informative title and axis labels, but challenge yourself to customize the features of

the figure further.

```
# Filter data for the year 2000
smalldat <- dat %>%
  filter(year == 2000)

# Create the histogram
gdppc_hist <- ggplot(smalldat, aes(x = WorldBank_gdppc_2010_con_estimate)) +
  geom_histogram(bins = 15,
                # Sets number of bins to 15
                color = "darkslategray",
                # Sets the color of the bin borders to dark slate gray
                fill = "cyan4") +
  # Fills the bins with a dark cyan color
  labs(title = "World GDP per Capita in 2000",
        x = "GDP per Capita (USD)",
        y = "Frequency") +
  # Add titles and labels to the plot
  coord_cartesian(ylim = c(0, 20),
                  xlim = c(4.75, 10.75))
  # Sets the limits of the x and y-axis

# Print the plot
print(gdppc_hist)
```



## Exercise 2: Kernel Density Plot of Land Orientation Over Time

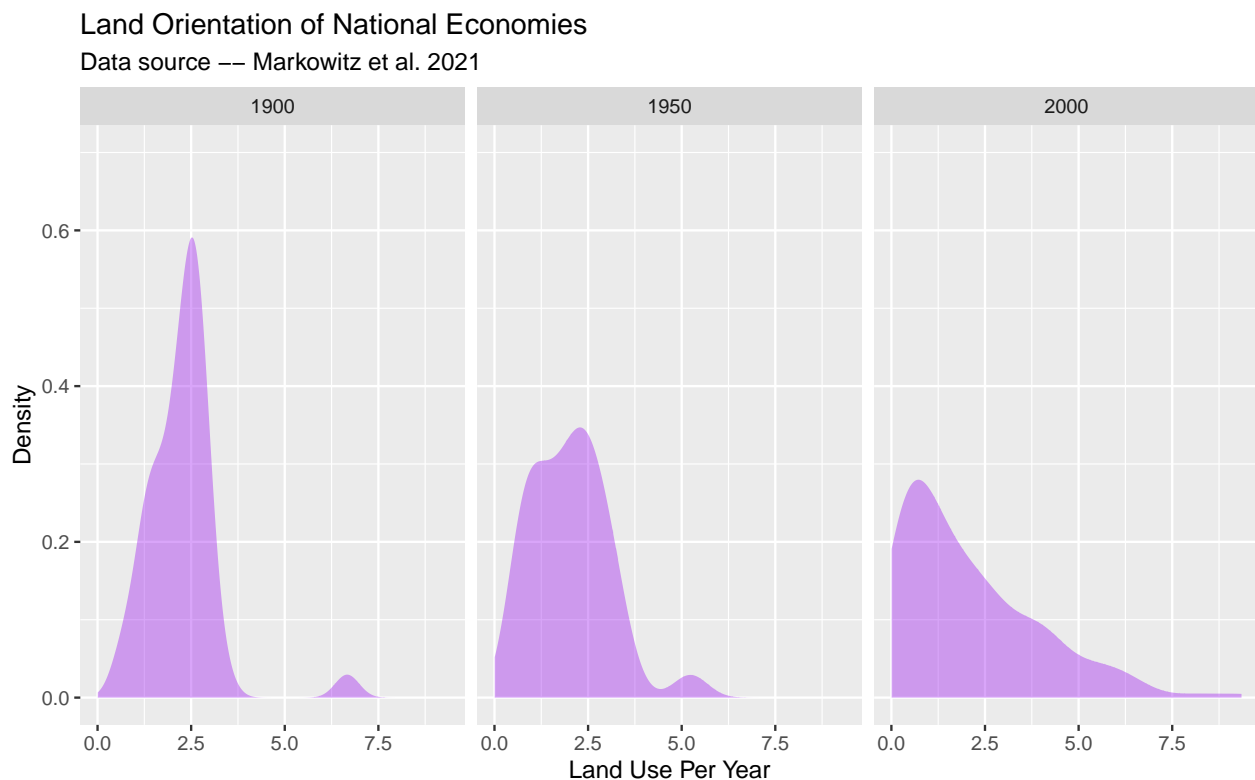
Now, create kernel density plots for land orientation for the years 1900, 1950, and 2000, and display them on the same graph to clearly show the trends over time. Once again, ensure you generate polished and

informative visualizations by customizing the aesthetic elements.

```
# Filter data for specified years
smalldat2 <- dat %>%
  filter(year %in% c(1900,1950,2000))

# Create the kernel density plot
land_kernel<- ggplot(smalldat2, aes(x = land_oriented_medium_continuous)) +
  geom_density(fill = "purple", alpha = 0.4, color = NA) +
  # Set fill color and transparency
  labs(title = "Land Orientation of National Economies",
        subtitle = "Data source -- Markowitz et al. 2021",
        x = "Land Use Per Year",
        y = "Density") +
  # Add titles and labels to the plot
  coord_cartesian(ylim = c(0, 0.7)) +
  # Sets the limits of the x and y-axis
  facet_wrap(~year)
  # Facet the plot by 'year' to show plots for each year in one graph

# Print the plot
print(land_kernel)
```



## Bonus Question

For an extra challenge, create a new visualization using a variable from the `ProductivePacifists_Data.RDATA` file that hasn't been visualized yet, or explore one of R's preloaded datasets (use `data()` to see the available options and `?dataset_name` to get details about the variables). Make sure to choose the visualization that best fits the data and its purpose—whether it's revealing trends, highlighting relationships, or comparing

distributions. Refer to the supplementary document on ‘What Makes a Good Visualization?’ (found in [SPEC Lab DV1 > Additional Materials](#)) for further references.

Experiment with `ggplot2`'s features to make a publication-quality visualization. Remember to apply everything you've learned so far about aesthetics, such as titles and axis labels. Challenge yourself to innovate by using features you haven't tried in previous assignments.