Q.1. Using the gapminder dataset for 2007:

a. Create three histograms of GDP per capita using different bin widths (1000, 5000, and 10000). Compare these visualizations and explain: a) How does the choice of bin width affect the appearance of the distribution? b) Which bin width provides the most meaningful insight into the data's structure?

b. Examine the population distribution (on log scale) in 2007 using density plots with Gaussian, Rectangular and Epanechnikov kernels. How do the three kernel functions affect the shape of the density estimation?

c. Create density plots of GDP per capita using three different bandwidths (1000, 5000, and 10000). How does bandwidth choice affect the smoothness and detail of the distribution?

d. Generate two visualizations to compare life expectancy distributions across continents. The first should use overlapping histograms, assigning a unique color to each continent. The second should employ faceted small multiples, with separate histograms for each continent.. Compare these visualizations and explain why the faceted approach is better.

e. Generate two visualizations for visualizing life expectancy distributions across continents. The first should use overlapping densities with transparency(alpha = 0.5), assigning a unique color to each continent. The second should employ ridge plots.(Use library 'ggridge'). Compare these visualizations and explain why the approach with ridge plots is better.

f. Using the gghalves library, create a half-violin, half-box plot visualization with points that effectively shows the GDP per capita distribution across continents.

Q.2. Using the economics dataset in ggplot2, complete the following tasks:

(a) Create a dual-axis plot where unemploy (number of unemployed individuals) is plotted on the primary y-axis and psavert (personal savings rate) is plotted on the secondary y-axis. You may need to rescale one of the variables to fit them on the same graph.

(b) Create an improved version using the secondary plot method(Use library 'patchwork'), where both variables are plotted separately but share a common x-axis (date).

(c) Explain why the second approach is preferable compared to the dual-axis plot.

Q.3. Using the mtcars library in ggplot2,

Creates a correlation matrix plot with scatter plots in the lower triangle, histogram distributions on the diagonal and correlation coefficients in the upper triangle using these 5 variables: mpg (Miles per gallon), disp (Displacement), hp (Horsepower), wt (Weight) and qsec (Quarter mile time).

(Hint: use ggpairs method in GGally library)

Q.4. Using the mpg dataset from the ggplot2 package, fit a multiple linear regression model to predict highway mileage (hwy) using engine displacement (displ), number of cylinders (cyl), and drivetrain type (drv). Then, extract the model coefficients along with their confidence intervals, excluding the intercept term. (Use library: 'broom')

Visualize the estimated coefficients from the regression model along with their confidence intervals using a point-range plot. The x-axis should represent the coefficient estimates, while the y-axis should list the predictor variables in reverse order. Additionally, include a vertical red reference line at x = 0 to indicate the threshold where a coefficient would have no effect.