Question 3 and 4

CJD

2025-11-21

Data

For referencing, please refer to the referencing Felix has completed, as I have taken his code for importing and cleaning the data, and have added my own code

```
# !!! DON"T FORGET TO CHANGE THE WORKING DIRECTORY TO YOUR OWN DIRECTORY !!!
setwd("/Users/charliedicker/Downloads/Big_Issue-main/Data")
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
library(tidyverse)
library(readr)
library(readxl)
library(patchwork)





# Importing Kenya Inequality
Ken_W_Ineq <- read_delim("Kenya_Wealth_Inequality.csv",
  delim = ";", escape_double = FALSE, col_names = FALSE,
  trim_ws = TRUE, skip = 1
) # This is why I would rather use xls file

Ken_I_Ineq <- read_delim("Kenya_Income_Inequality.csv",
  delim = ";", escape_double = FALSE, col_names = FALSE,
  trim_ws = TRUE, skip = 1
)

# Importing Kenya Income Growth
Ken_I_Growth <- read_excel("Bang_Income.xlsx")

# Importing Bangladesh Inequality
Bang_W_Ineq <- read_excel("Bang_Wealth_Inequality.xlsx", col_names = FALSE)
Bang_I_Ineq <- read_excel("Bang_Income_Inequality.xlsx", col_names = FALSE)

# Importing Bangladesh Income Growth
Bang_I_Growth <- read_excel("Kenya_Income.xlsx")

# Importing Kenya Trade
Trade_df <- read_excel("Bang_and_Ken_Trade.xlsx", col_names = FALSE)
```

```r
# Importing Bangladesh Trade
```

```r
# Make a function for cleaning data sets
clean_data_inequality <- function(x) {
  colnames(x) <- c("Country", "Indicator", "Percentile", "Year", "Value")
  # Columns of Inequality data sets are all in this order, check when use for others
  x <- x %>%
    pivot_wider(
      names_from = Percentile,
      values_from = Value
    ) %>%
    filter(!if_all(c(pall, p0p50, p50p90, p90p100, p99p100), is.na)) %>%
    # filter out those rows where all the values are NA
    select(Country, Year, pall, p0p50, p50p90, p90p100, p99p100) %>%
    # To ensure the columns are in correct order and delete indicator column
    group_by(Country, Year) %>%
    summarise(
      across(
        c(pall, p0p50, p50p90, p90p100, p99p100),
        ~ first(na.omit(.))
      ),
      .groups = "drop"
    )
  # Note that previously we have 5 lines for a single year, and each
  # line only shows a single indicator. By doing this, we combine the data together.
  colnames(x) <- c(
    "Country", "Year", "Gini_Coeff", "Share_Bottom50",
    "Share_Middle40", "Share_Top10", "Share_Top1"
  )
  return(x)
}

#I added this function
clean_data_growth <- function(x) {
  colnames(x) <- c("Year", "National Income")
  x <- x %>% mutate("Growth" = ((`National Income` - lag(`National Income`))/ lag(`National Income`) *
  filter(!is.na(Growth))

  return(x)
}

# I added this function to clean trade data
clean_trade <- function(x) {
  colnames(x) <- c("Percentile","Year", "Bang_Trade", "Ken_Trade")
  x$Percentile <- NULL
  x <- x %>% filter(!is.na(Bang_Trade) & !is.na(Ken_Trade))
  x$Bang_Trade <- as.numeric(x$Bang_Trade)
  x$Ken_Trade <- as.numeric(x$Ken_Trade)


  return(x)
}
```

```r
Bang_I_Ineq_wider <- clean_data_inequality(Bang_I_Ineq)
Bang_W_Ineq_wider <- clean_data_inequality(Bang_W_Ineq)

#I added this
Bang_I_Growth_cleaned <- clean_data_growth(Bang_I_Growth)


Ken_I_Ineq_wider <- clean_data_inequality(Ken_I_Ineq)
Ken_W_Ineq_wider <- clean_data_inequality(Ken_W_Ineq)

# I added this
Ken_I_Growth_cleaned <- clean_data_growth(Ken_I_Growth)

# I added this
Trade_cleaned <- clean_trade(Trade_df)
```

```
## Warning in clean_trade(Trade_df): NAs introduced by coercion
## Warning in clean_trade(Trade_df): NAs introduced by coercion
```

```r
Trade_cleaned <- Trade_cleaned[-1, ]


longer_format <- function(x) {
  x %>% pivot_longer(
    cols = c(
      Gini_Coeff, Share_Bottom50, Share_Middle40,
      Share_Top10, Share_Top1
    ),
    names_to = "Indicator",
    values_to = "Value"
  )
}

Bang_I_Ineq_longer <- longer_format(Bang_I_Ineq_wider)
Bang_W_Ineq_longer <- longer_format(Bang_W_Ineq_wider)
Ken_I_Ineq_longer <- longer_format(Ken_I_Ineq_wider)
Ken_W_Ineq_longer <- longer_format(Ken_W_Ineq_wider)
```

Question 3

First, we need to join inequality and growth data. I have written a quick function, using inner_join in order to remove any years with only 1 value

```r
# Function to join Inequality and Growth data
to_join <- function(x, y) {
  x <- x %>%
    inner_join(y, by = "Year")
  return(x)
}
```

I then wrote 3 functions, to calculate correlation between the Gini Coefficient and National Income Growth. I used Pearson Correlation Coefficient with and without outliers, and the Spearman Correlation Coefficient with outliers.

I cleaned the data and got rid of outliers using the method of anything which is less than the 1st quartile - (1.5 x the interquartile range), or greater than the 3rd quartile + (1.5 x the interquartile range). This is a simple and common way to remove outliers.

GDP growth often contains extreme but economically relevant fluctuations, while Gini coefficients vary within a narrower range (between 0 and 1). Therefore, I removed outliers only from the growth variable to provide a robustness check, but retain inequality values in full. This approach prevents the loss of important economic information while allowing me to assess whether extreme growth shocks drive the correlation.

```r
# Function calculates Pearson Correlation Coeficcient (with outliers)
calc_correlation <- function(x) {
  return(cor(x$Gini_Coeff, x$Growth))
}

# Function calculates Pearson Correlation Coefficient (without outliers) (I used IQR)
calc_correlation_noOutliers <- function(x) {
  Q1 <- quantile(x$Growth, 0.25, na.rm = TRUE)
  Q3 <- quantile(x$Growth, 0.75, na.rm = TRUE)
  IQR <- Q3 - Q1

  y <- x %>%
  filter(Growth > Q1 - 1.5*IQR & Growth < Q3 + 1.5*IQR)

  return(cor(y$Gini_Coeff, y$Growth))
}

# Function calculates Spearman Correlation Coefficient (with outliers)
  calc_correlation_spearman <- function(x){
  return(cor(x$Gini_Coeff, x$Growth, method = "spearman"))
}
```

I then created 2 functions. One to create a graph with the outliers, and one to create a graph without outliers. This allows me to view if outliers are driving correlation, or vice versa, if there is a higher correlations without volatile fluctuations in national income growth.

```r
# Function to create the graphs (with outliers)
graph <- function(x, y) {
 return(ggplot(x, aes(x = Growth, y = Gini_Coeff, colour = Type)) +
  geom_point(size = 0.4, alpha = 0.7) +
  geom_smooth(method = "lm", size = 0.2) +
  theme_minimal(base_size = 6) +
  theme(
    panel.background = element_rect(fill = "#f7f7f7", color = NA),
    panel.grid.major = element_line(color = "#dcdcdc"),
    panel.grid.minor = element_line(color = "#eaeaea"),
    axis.title = element_text(color = "#333333", face = "bold"),
    axis.text = element_text(color = "#333333"),
    plot.title.position = "plot",
    plot.title = element_text(
    lineheight = 1.1,
    margin = margin(b = 10, t = 5),
```

```
    hjust = 0.5 )) +
  labs(
    title = y,
    x = "Growth (%)",
    y = "Gini Coefficient",
    colour = "Type"
  )
  )
}

# Function to create the graphs (without outliers)
graph_noOutliers <- function(x, y) {
  Q1 <- quantile(x$Growth, 0.25, na.rm = TRUE)
  Q3 <- quantile(x$Growth, 0.75, na.rm = TRUE)
  IQR <- Q3 - Q1

  z <- x %>%
  filter(Growth > Q1 - 1.5*IQR & Growth < Q3 + 1.5*IQR)

 return(ggplot(z, aes(x = Growth, y = Gini_Coeff, colour = Type)) +
  geom_point( size = 0.4, alpha = 0.7) +
  geom_smooth(method = "lm", size = 0.2) +
  theme_minimal(base_size = 6) +
  theme(
    panel.background = element_rect(fill = "#f7f7f7", color = NA),
    panel.grid.major = element_line(color = "#dcdcdc"),
    panel.grid.minor = element_line(color = "#eaeaea"),
    axis.title = element_text(color = "#333333", face = "bold"),
    axis.text = element_text(color = "#333333"),
    plot.title.position = "plot",
    plot.title = element_text(
    lineheight = 1.1,
    margin = margin(b = 10, t = 5),
    hjust = 0.5 )) +
  labs(
    title = stringr::str_wrap(y, width = 40),
    x = "Growth (%)",
    y = "Gini Coefficient",
    colour = "Type"
  ))
}
```

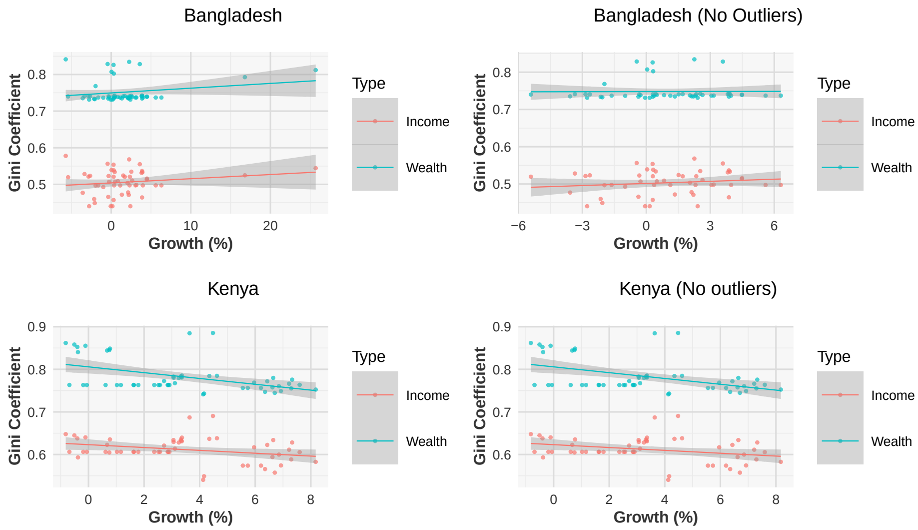
This next section is just passing data into the functions, and getting returned values or graphs.

```
# Joining both inequality data and income growth by the year
Bang_I_Ineq_and_Growth <- merge(Bang_I_Growth_cleaned, Bang_I_Ineq_wider, by = "Year")
Bang_W_Ineq_and_Growth <- merge(Bang_I_Growth_cleaned, Bang_W_Ineq_wider, by = "Year")
Ken_I_Ineq_and_Growth <- merge(Ken_I_Growth_cleaned, Ken_I_Ineq_wider, by = "Year")
Ken_W_Ineq_and_Growth <- merge(Ken_I_Growth_cleaned, Ken_W_Ineq_wider, by = "Year")

Bang_I_Ineq_and_Growth$Type <- "Income"
Bang_W_Ineq_and_Growth$Type <- "Wealth"
```

```r
Ken_I_Ineq_and_Growth$Type <- "Income"
Ken_W_Ineq_and_Growth$Type <- "Wealth"

Bang_Ineq_and_Growth <- rbind(Bang_I_Ineq_and_Growth, Bang_W_Ineq_and_Growth)
Ken_Ineq_and_Growth <- rbind(Ken_I_Ineq_and_Growth, Ken_W_Ineq_and_Growth)



# Calculating the pearson correlation coefficient with outliers
Bang_I_Ineq_and_Growth_Correlation <- calc_correlation(Bang_I_Ineq_and_Growth)
Bang_W_Ineq_and_Growth_Correlation <- calc_correlation(Bang_W_Ineq_and_Growth)
Ken_I_Ineq_and_Growth_Correlation <- calc_correlation(Ken_I_Ineq_and_Growth)
Ken_W_Ineq_and_Growth_Correlation <- calc_correlation(Ken_W_Ineq_and_Growth)

# Calculating the pearson correlation coefficient without outliers
Bang_I_Ineq_and_Growth_Correlation_NoOutliers <- calc_correlation_noOutliers(Bang_I_Ineq_and_Growth)
Bang_W_Ineq_and_Growth_Correlation_NoOutliers <- calc_correlation_noOutliers(Bang_W_Ineq_and_Growth)
Ken_I_Ineq_and_Growth_Correlation_NoOutliers <- calc_correlation_noOutliers(Ken_I_Ineq_and_Growth)
Ken_W_Ineq_and_Growth_Correlation_NoOutliers <- calc_correlation_noOutliers(Ken_W_Ineq_and_Growth)

# Calculating the spearman correlation coefficient with outliers
Bang_I_Ineq_and_Growth_Correlation_Spearman <- calc_correlation_spearman(Bang_I_Ineq_and_Growth)
Bang_W_Ineq_and_Growth_Correlation_Spearman <- calc_correlation_spearman(Bang_W_Ineq_and_Growth)
Ken_I_Ineq_and_Growth_Correlation_Spearman <- calc_correlation_spearman(Ken_I_Ineq_and_Growth)
Ken_W_Ineq_and_Growth_Correlation_Spearman <- calc_correlation_spearman(Ken_W_Ineq_and_Growth)



# Plotting the graphs
(graph(Bang_Ineq_and_Growth, "Bangladesh") |
graph_noOutliers(Bang_Ineq_and_Growth, "Bangladesh (No Outliers)")) /
(graph(Ken_Ineq_and_Growth, "Kenya") |
graph_noOutliers(Ken_Ineq_and_Growth, "Kenya (No outliers)"))
```

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



We then calculate the p value, with a hypothesis test of: H0: p = 0 H1: p != 0

With a significance level of 0.05

```
cor_p_value <- function(r, n) {
  t_value <- r * sqrt((n - 2) / (1 - r^2))
  degf <- n - 2
  p_value <- 2 * (1 - pt(abs(t_value), degf))
  p_value_char <- as.character(p_value)
  if (p_value < 0.05) {
    return(paste(p_value_char, "There is sufficient evidence showing correlation"))
  }
  else
  {
    return(paste(p_value_char, "There is insufficient evidence to show correlation"))
  }

}

# Function to create a dataframe with no outliers
no_outliers <- function(x) {
  Q1 <- quantile(x$Growth, 0.25, na.rm = TRUE)
  Q3 <- quantile(x$Growth, 0.75, na.rm = TRUE)
  IQR <- Q3 - Q1

  y <- x %>%
  filter(Growth > Q1 - 1.5*IQR & Growth < Q3 + 1.5*IQR)

  return(y)
}


Bang_I_Ineq_and_Growth_Correlation_Pearson_Sig = cor_p_value(Bang_I_Ineq_and_Growth_Correlation, nrow(Ba
Bang_I_Ineq_and_Growth_Correlation_Pearson_noOutliers_Sig = cor_p_value(Bang_I_Ineq_and_Growth_Correlati
Bang_I_Ineq_and_Growth_Correlation_Spearman_Sig = cor_p_value(Bang_I_Ineq_and_Growth_Correlation_Spearma
```

```
Bang_W_Ineq_and_Growth_Correlation_Pearson_Sig = cor_p_value(Bang_W_Ineq_and_Growth_Correlation, nrow(Ba
Bang_W_Ineq_and_Growth_Correlation_Pearson_noOutliers_Sig = cor_p_value(Bang_W_Ineq_and_Growth_Correlati
Bang_W_Ineq_and_Growth_Correlation_Spearman_Sig = cor_p_value(Bang_W_Ineq_and_Growth_Correlation_Spearma

Ken_I_Ineq_and_Growth_Correlation_Pearson_Sig = cor_p_value(Ken_I_Ineq_and_Growth_Correlation, nrow(Ken_
Ken_I_Ineq_and_Growth_Correlation_Pearson_noOutliers_Sig = cor_p_value(Ken_I_Ineq_and_Growth_Correlatio
Ken_I_Ineq_and_Growth_Correlation_Spearman_Sig = cor_p_value(Ken_I_Ineq_and_Growth_Correlation_Spearman

Ken_W_Ineq_and_Growth_Correlation_Pearson_Sig = cor_p_value(Ken_W_Ineq_and_Growth_Correlation, nrow(Ken_
Ken_W_Ineq_and_Growth_Correlation_Pearson_noOutliers_Sig = cor_p_value(Ken_W_Ineq_and_Growth_Correlatio
Ken_W_Ineq_and_Growth_Correlation_Spearman_Sig = cor_p_value(Ken_W_Ineq_and_Growth_Correlation_Spearman
```

This next chunk is taking all of the correlation values I have calculated and created a data frame out of it

```
correlation_df <- data.frame(
  Country = c("Bangladesh Income Pearson", "Bangladesh Wealth Pearson", "Kenya Income Pearson", "Kenya W

  Correlation = c(
    Bang_I_Ineq_and_Growth_Correlation,
    Bang_W_Ineq_and_Growth_Correlation,
    Ken_I_Ineq_and_Growth_Correlation,
    Ken_W_Ineq_and_Growth_Correlation,
    Bang_I_Ineq_and_Growth_Correlation_NoOutliers,
    Bang_W_Ineq_and_Growth_Correlation_NoOutliers,
    Ken_I_Ineq_and_Growth_Correlation_NoOutliers,
    Ken_W_Ineq_and_Growth_Correlation_NoOutliers,
    Bang_I_Ineq_and_Growth_Correlation_Spearman,
    Bang_W_Ineq_and_Growth_Correlation_Spearman,
    Ken_I_Ineq_and_Growth_Correlation_Spearman,
    Ken_W_Ineq_and_Growth_Correlation_Spearman
  ),

  P_Value = c(
  Bang_I_Ineq_and_Growth_Correlation_Pearson_Sig,
  Bang_I_Ineq_and_Growth_Correlation_Pearson_noOutliers_Sig,
  Bang_I_Ineq_and_Growth_Correlation_Spearman_Sig,

  Bang_W_Ineq_and_Growth_Correlation_Pearson_Sig,
  Bang_W_Ineq_and_Growth_Correlation_Pearson_noOutliers_Sig,
  Bang_W_Ineq_and_Growth_Correlation_Spearman_Sig,

  Ken_I_Ineq_and_Growth_Correlation_Pearson_Sig,
  Ken_I_Ineq_and_Growth_Correlation_Pearson_noOutliers_Sig,
  Ken_I_Ineq_and_Growth_Correlation_Spearman_Sig,

  Ken_W_Ineq_and_Growth_Correlation_Pearson_Sig,
  Ken_W_Ineq_and_Growth_Correlation_Pearson_noOutliers_Sig,
  Ken_W_Ineq_and_Growth_Correlation_Spearman_Sig
))

correlation_df
```

```
##                                      Country  Correlation
```

```
## 1                Bangladesh Income Pearson  0.161933870
## 2                Bangladesh Wealth Pearson  0.196492884
## 3                     Kenya Income Pearson -0.300089929
## 4                     Kenya Wealth Pearson -0.463112553
## 5  Bangladesh Income Pearson (no outliers)  0.186709229
## 6  Bangladesh Wealth Pearson (no outliers) -0.006729303
## 7       Kenya Income Pearson (no outliers) -0.300089929
## 8       Kenya Wealth Pearson (no outliers) -0.463112553
## 9               Bangladesh Income Spearman  0.139569609
## 10              Bangladesh Wealth Spearman  0.111406254
## 11                   Kenya Income Spearman -0.285467035
## 12                   Kenya Wealth Spearman -0.378578910
##                                                                  P_Value
## 1   0.242056224902433 There is insufficient evidence to show correlation
## 2    0.19418695207814 There is insufficient evidence to show correlation
## 3   0.314145323540167 There is insufficient evidence to show correlation
## 4   0.154428807283367 There is insufficient evidence to show correlation
## 5    0.96300711595396 There is insufficient evidence to show correlation
## 6   0.422546753733279 There is insufficient evidence to show correlation
## 7    0.0274759172689045 There is sufficient evidence showing correlation
## 8    0.0274759172689045 There is sufficient evidence showing correlation
## 9    0.0364010904097185 There is sufficient evidence showing correlation
## 10 0.000421328413568789 There is sufficient evidence showing correlation
## 11 0.000421328413568789 There is sufficient evidence showing correlation
## 12   0.0047609695818327 There is sufficient evidence showing correlation
```

Analysis on the data:

Sufficient evidence for Kenya Income and Wealth Pearson (no outliers) and Spearman, Bangladesh Income and Wealth Spearman

Question 4

Use R to analyse the correlation between international trade and income inequality, and international trade and income growth for the two countries. [15 marks]

Chunk with functions calculate dfs with no outliers

```
calc_outliers_ <- function(x) {
  Q1 <- quantile(x, 0.25, na.rm = TRUE)
  Q3 <- quantile(x, 0.75, na.rm = TRUE)
  IQR <- Q3 - Q1

  x[x < Q1 - 1.5*IQR | x > Q3 + 1.5*IQR] <- NA
  return(x)
}
```

Chunk to pass in all previous data

```
# Merging the dfs into 2 large dfs
All_Data_Bang <- merge(Trade_cleaned, Bang_I_Ineq_and_Growth, by = "Year")
All_Data_Ken <- merge(Trade_cleaned, Ken_I_Ineq_and_Growth, by = "Year")
All_Data_Bang$Bang_Trade <- as.numeric(All_Data_Bang$Bang_Trade)
```

```
All_Data_Ken$Ken_Trade <- as.numeric(All_Data_Ken$Ken_Trade)

# Creating correlation data
Bang_Trade_Ineq_Corr <- cor(All_Data_Bang$Bang_Trade, All_Data_Bang$Gini_Coeff)
Bang_Trade_Growth_Corr <- cor(All_Data_Bang$Bang_Trade, All_Data_Bang$Growth)
Ken_Trade_Ineq_Corr <- cor(All_Data_Ken$Ken_Trade, All_Data_Ken$Gini_Coeff)
Ken_Trade_Growth_Corr <- cor(All_Data_Ken$Ken_Trade, All_Data_Ken$Growth)




# No outliers
All_Data_Bang$Bang_Trade_clean <- calc_outliers_(All_Data_Bang$Bang_Trade)
All_Data_Bang$Gini_Coeff_clean <- calc_outliers_(All_Data_Bang$Gini_Coeff)
Bang_clean <- na.omit(All_Data_Bang[, c("Bang_Trade_clean", "Gini_Coeff_clean")])
Bang_Trade_Ineq_Corr_noOutliers <- cor(Bang_clean$Bang_Trade_clean, Bang_clean$Gini_Coeff_clean)

All_Data_Bang$Growth_clean <- calc_outliers_(All_Data_Bang$Growth)
Bang_clean <- na.omit(All_Data_Bang[, c("Bang_Trade_clean", "Growth_clean")])
Bang_Trade_Growth_Corr_noOutliers <- cor(Bang_clean$Bang_Trade_clean, Bang_clean$Growth_clean)

All_Data_Ken$Ken_Trade_clean <- calc_outliers_(All_Data_Ken$Ken_Trade)
All_Data_Ken$Gini_Coeff_clean <- calc_outliers_(All_Data_Ken$Gini_Coeff)
Ken_clean <- na.omit(All_Data_Ken[, c("Ken_Trade_clean", "Gini_Coeff_clean")])
Ken_Trade_Ineq_Corr_noOutliers <- cor(Ken_clean$Ken_Trade_clean, Ken_clean$Gini_Coeff_clean)

All_Data_Ken$Growth_clean <- calc_outliers_(All_Data_Ken$Growth)
Ken_clean <- na.omit(All_Data_Ken[, c("Ken_Trade_clean", "Growth_clean")])
Ken_Trade_Growth_Corr_noOutliers <- cor(Ken_clean$Ken_Trade_clean, Ken_clean$Growth_clean)

colnames(All_Data_Bang)[colnames(All_Data_Bang) == "Bang_Trade_clean"] <- "Trade_clean"
colnames(All_Data_Ken)[colnames(All_Data_Ken) == "Ken_Trade_clean"] <- "Trade_clean"

All_Data <- rbind(All_Data_Bang, All_Data_Ken)




# Spearman correlations
Bang_Trade_Ineq_Corr_spearman <- cor(All_Data_Bang$Bang_Trade, All_Data_Bang$Gini_Coeff, method = "spear
Bang_Trade_Growth_Corr_spearman <- cor(All_Data_Bang$Bang_Trade, All_Data_Bang$Growth, method = "spearma
Ken_Trade_Ineq_Corr_spearman <- cor(All_Data_Ken$Ken_Trade, All_Data_Ken$Gini_Coeff, method = "spearman
Ken_Trade_Growth_Corr_spearman <- cor(All_Data_Ken$Ken_Trade, All_Data_Ken$Growth, method = "spearman")
```

Chunk with functions to display the graph

```
# Graphing Trade vs Gini
(ggplot() +

  # Bangladesh trade
  geom_point(data = All_Data,
```

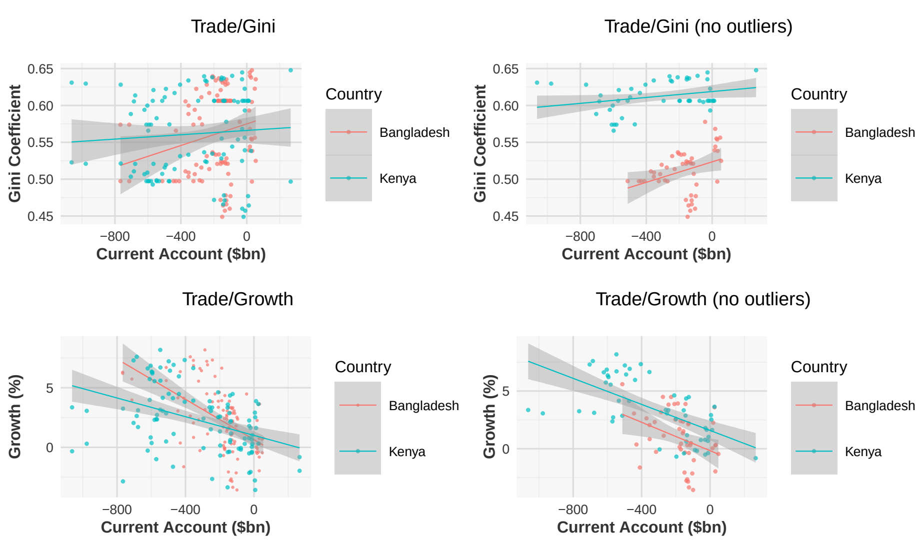
```r
             aes(x = Bang_Trade,
                 y = Gini_Coeff_clean,
                 colour = "Bangladesh"),
             size = 0.4, alpha = 0.7) +
  geom_smooth(data = All_Data,
              aes(x = Bang_Trade,
                  y = Gini_Coeff_clean,
                  colour = "Bangladesh"),
              method = "lm", size = 0.2) +

  # Kenya trade
  geom_point(data = All_Data,
             aes(x = Ken_Trade,
                 y = Gini_Coeff_clean,
                 colour = "Kenya"),
             size = 0.4, alpha = 0.7) +
  geom_smooth(data = All_Data,
              aes(x = Ken_Trade,
                  y = Gini_Coeff_clean,
                  colour = "Kenya"),
              method = "lm", size = 0.2) +

  # Theme
  theme_minimal(base_size = 6) +
  theme(
    panel.background = element_rect(fill = "#f7f7f7", color = NA),
    panel.grid.major = element_line(color = "#dcdcdc"),
    panel.grid.minor = element_line(color = "#eaeaea"),
    axis.title = element_text(color = "#333333", face = "bold"),
    axis.text = element_text(color = "#333333"),
    plot.title.position = "plot",
    plot.title = element_text(
      lineheight = 1.1,
      margin = margin(b = 10, t = 5),
      hjust = 0.5
    )
  ) +

  # Labels
  labs(
    title = "Trade/Gini",
    x = "Current Account ($bn)",
    y = "Gini Coefficient",
    colour = "Country"
  ) |
(ggplot(All_Data, aes(x = Trade_clean, y = Gini_Coeff_clean, colour = Country)) +
  geom_point(size = 0.4, alpha = 0.7) +
  geom_smooth(method = "lm", size = 0.2) +
  theme_minimal(base_size = 6) +
  theme(
    panel.background = element_rect(fill = "#f7f7f7", color = NA),
    panel.grid.major = element_line(color = "#dcdcdc"),
    panel.grid.minor = element_line(color = "#eaeaea"),
```

```
    axis.title = element_text(color = "#333333", face = "bold"),
    axis.text = element_text(color = "#333333"),
    plot.title.position = "plot",
    plot.title = element_text(
    lineheight = 1.1,
    margin = margin(b = 10, t = 5),
    hjust = 0.5 )) +
  labs(
    title = "Trade/Gini (no outliers)",
    x = "Current Account ($bn)",
    y = "Gini Coefficient",
    colour = "Country"
  ))) /
(ggplot() +

  # Bangladesh trade
  geom_point(data = All_Data,
             aes(x = Bang_Trade,
                 y = Growth_clean,
                 colour = "Bangladesh"),
             size = 0.2, alpha = 0.7) +
  geom_smooth(data = All_Data,
              aes(x = Bang_Trade,
                  y = Growth_clean,
                  colour = "Bangladesh"),
              method = "lm", size = 0.2) +

  # Kenya trade
  geom_point(data = All_Data,
             aes(x = Ken_Trade,
                 y = Growth_clean,
                 colour = "Kenya"),
             size = 0.4, alpha = 0.7) +
  geom_smooth(data = All_Data,
              aes(x = Ken_Trade,
                  y = Growth_clean,
                  colour = "Kenya"),
              method = "lm", size = 0.2) +

  # Theme
  theme_minimal(base_size = 6) +
  theme(
    panel.background = element_rect(fill = "#f7f7f7", color = NA),
    panel.grid.major = element_line(color = "#dcdcdc"),
    panel.grid.minor = element_line(color = "#eaeaea"),
    axis.title = element_text(color = "#333333", face = "bold"),
    axis.text = element_text(color = "#333333"),
    plot.title.position = "plot",
    plot.title = element_text(
      lineheight = 1.1,
      margin = margin(b = 10, t = 5),
      hjust = 0.5
    )
```

```
  ) +

  # Labels
  labs(
    title = "Trade/Growth",
    x = "Current Account ($bn)",
    y = "Growth (%)",
    colour = "Country"
  ) |
ggplot(All_Data, aes(x = Trade_clean, y = Growth_clean, colour = Country)) +
  geom_point(size = 0.4, alpha = 0.7) +
  geom_smooth(method = "lm", size = 0.2) +
  theme_minimal(base_size = 6) +
  theme(
    panel.background = element_rect(fill = "#f7f7f7", color = NA),
    panel.grid.major = element_line(color = "#dcdcdc"),
    panel.grid.minor = element_line(color = "#eaeaea"),
    axis.title = element_text(color = "#333333", face = "bold"),
    axis.text = element_text(color = "#333333"),
    plot.title.position = "plot",
    plot.title = element_text(
    lineheight = 1.1,
    margin = margin(b = 10, t = 5),
    hjust = 0.5 )) +
  labs(
    title = "Trade/Growth (no outliers)",
    x = "Current Account ($bn)",
    y = "Growth (%)",
    colour = "Country"
  ))
```

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

## Warning: Removed 9 rows containing non-finite outside the scale range
## (`stat_smooth()`).

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

## Warning: Removed 9 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 9 rows containing missing values or values outside the scale range
## (`geom_point()`).
## Removed 9 rows containing missing values or values outside the scale range
## (`geom_point()`).

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

## Warning: Removed 11 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 11 rows containing missing values or values outside the scale range
## (`geom_point()`).
```

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

## Warning: Removed 4 rows containing non-finite outside the scale range
## (`stat_smooth()`).

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

## Warning: Removed 4 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).
## Removed 4 rows containing missing values or values outside the scale range
## (`geom_point()`).

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

## Warning: Removed 6 rows containing non-finite outside the scale range
## (`stat_smooth()`).

## Warning: Removed 6 rows containing missing values or values outside the scale range
## (`geom_point()`).
```



We then calculate the p value, with a hypothesis test of: H0: p = 0 H1: p != 0

With a significance level of 0.05

```r
# Function to calculate the p value
p_value_calc <- function(r, n)
{

  t <- r * sqrt(n - 2) / sqrt(1 - r^2)
  df <- n - 2
```

```r
  p <- 2 * (1 - pt(abs(t), df))
    p_char <- as.character(p)

  if(p < 0.05) {
    return(paste(p_char, "There is sufficient evidence showing correlation"))
  }
  else {
    return(paste(p_char, "There is insufficient evidence to showing correlation"))
  }
}

Bang_Trade_Ineq_p <- p_value_calc(Bang_Trade_Ineq_Corr, nrow(All_Data_Bang))
Bang_Trade_Growth_p <- p_value_calc(Bang_Trade_Growth_Corr, nrow(All_Data_Bang))
Ken_Trade_Ineq_p <- p_value_calc(Ken_Trade_Ineq_Corr, nrow(All_Data_Ken))
Ken_Trade_Growth_p <- p_value_calc(Ken_Trade_Growth_Corr, nrow(All_Data_Ken))


All_Data_Bang_clean <- All_Data_Bang[, c("Bang_Trade", "Gini_Coeff")]
All_Data_Bang_clean <- na.omit(All_Data_Bang_clean)

All_Data_Ken_clean <- All_Data_Ken[, c("Ken_Trade", "Gini_Coeff")]
All_Data_Ken_clean <- na.omit(All_Data_Ken_clean)

Bang_Trade_Ineq_p_noOutliers <- p_value_calc(Bang_Trade_Ineq_Corr_noOutliers, nrow(All_Data_Bang_clean)
Bang_Trade_Growth_p_noOutliers <- p_value_calc(Bang_Trade_Growth_Corr_noOutliers, nrow(All_Data_Bang_cle
Ken_Trade_Ineq_p_noOutliers <- p_value_calc(Ken_Trade_Ineq_Corr_noOutliers, nrow(All_Data_Ken_clean))
Ken_Trade_Growth_p_noOutliers <- p_value_calc(Ken_Trade_Growth_Corr_noOutliers, nrow(All_Data_Ken_clean

Bang_Trade_Ineq_p_spearman <- p_value_calc(Bang_Trade_Ineq_Corr_spearman, nrow(All_Data_Bang))
Bang_Trade_Growth_p_spearman <- p_value_calc(Bang_Trade_Growth_Corr_spearman, nrow(All_Data_Bang))
Ken_Trade_Ineq_p_spearman <- p_value_calc(Ken_Trade_Ineq_Corr_spearman, nrow(All_Data_Ken))
Ken_Trade_Growth_p_spearman <- p_value_calc(Ken_Trade_Growth_Corr_spearman, nrow(All_Data_Ken))



correlation_df_4 <- data.frame(
  Country_Trade_Comp = c("Bangladesh Inequality Pearson", "Bangladesh Growth Pearson", "Kenya Inequalit

  Correlation_ = c(
  Bang_Trade_Ineq_Corr,
  Bang_Trade_Growth_Corr,
  Ken_Trade_Ineq_Corr,
  Ken_Trade_Growth_Corr,
  Bang_Trade_Ineq_Corr_noOutliers,
  Bang_Trade_Growth_Corr_noOutliers,
  Ken_Trade_Ineq_Corr_noOutliers,
  Ken_Trade_Growth_Corr_noOutliers,
  Bang_Trade_Ineq_Corr_spearman,
  Bang_Trade_Growth_Corr_spearman,
  Ken_Trade_Ineq_Corr_spearman,
  Ken_Trade_Growth_Corr_spearman

  ),
```

```
  P_Value_ = c(
    Bang_Trade_Ineq_p,
    Bang_Trade_Growth_p,
    Ken_Trade_Ineq_p,
    Ken_Trade_Growth_p,
    Bang_Trade_Ineq_p_noOutliers,
    Bang_Trade_Growth_p_noOutliers,
    Ken_Trade_Ineq_p_noOutliers,
    Ken_Trade_Growth_p_noOutliers,
    Bang_Trade_Ineq_p_spearman,
    Bang_Trade_Growth_p_spearman,
    Ken_Trade_Ineq_p_spearman,
    Ken_Trade_Growth_p_spearman
))

correlation_df_4
```

```
##                             Country_Trade_Comp Correlation_
## 1               Bangladesh Inequality Pearson   0.24152786
## 2                   Bangladesh Growth Pearson  -0.02158368
## 3                    Kenya Inequality Pearson   0.36294178
## 4                        Kenya Growth Pearson  -0.62513074
## 5  Bangladesh Inequality Pearson (no outliers)   0.33389853
## 6      Bangladesh Growth Pearson (no outliers)  -0.35802412
## 7       Kenya Inequality Pearson (no outliers)   0.28321365
## 8           Kenya Growth Pearson (no outliers)  -0.62513074
## 9              Bangladesh Inequality Spearman   0.28494031
## 10                 Bangladesh Growth Spearman  -0.34903755
## 11                  Kenya Inequality Spearman   0.35124622
## 12                      Kenya Growth Spearman  -0.63765962
##                                                                     P_Value_
## 1  0.0784956098191993 There is insufficient evidence to showing correlation
## 2   0.876889318374959 There is insufficient evidence to showing correlation
## 3      0.00699019520730992 There is sufficient evidence showing correlation
## 4     4.34593126596994e-07 There is sufficient evidence showing correlation
## 5       0.0136060787720877 There is sufficient evidence showing correlation
## 6      0.00785795239011255 There is sufficient evidence showing correlation
## 7       0.0379691859782416 There is sufficient evidence showing correlation
## 8     4.34593126596994e-07 There is sufficient evidence showing correlation
## 9        0.036762748462349 There is sufficient evidence showing correlation
## 10     0.00968746866326264 There is sufficient evidence showing correlation
## 11     0.00920666913912882 There is sufficient evidence showing correlation
## 12    2.15304707396058e-07 There is sufficient evidence showing correlation
```

Analysis:

#AI Usage I have used AI in producing my graphs, specifically, for making the graphs have a nice theme. The prompt used: using ggplot2, write me code to change the colour of the points, of the line, and change the background to be nice