A Comparative Analysis of Employment Services Across Countries and Regions with Factors Affecting it Using R Studio and Power Bi

Table of Contents

Part One: Introduction, Data Preparation and Cleaning

1.0 Introduction

Global economic disparities, particularly in employment, are critical facets of the socio-economic landscape, the analysis of employment services across different regions within a country has emerged as a critical area of study. The disparities in employment opportunities, wage structures, and overall job satisfaction among various regions often underscore the complex interplay of economic, social, and geographic factors. This research endeavours to conduct a comprehensive comparative analysis of employment services across diverse regions within a country, with a specific focus on identifying the factors that contribute to the observed disparities.

The defined problem at the heart of this study is the persistent and, at times, widening gap in employment services among different regions. Research has indicated that regional imbalances in employment opportunities can have profound consequences on individual well-being, economic growth, and societal harmony (Smith et al., 2019; Gupta & Sharma, 2021). Understanding the nuanced factors that contribute to these disparities is crucial for policymakers, businesses, and community leaders aiming to formulate effective strategies for promoting equitable employment distribution.

The study is grounded in the premise that understanding the intricacies of employment dynamics is pivotal for informed policy-making and international development efforts. The main idea is that understanding the complicated nature of employment is vital for creating smart policies and supporting successful global development projects. To inform the exploration of this research, studies such as the work of Salas-Velasco and López-Bazo (2020) on regional wage differentials and the research by Jones et al. (2018) on the impact of urbanization on employment opportunities will serve as foundational references

1.1 Research Question

Based on economic trends, factors and indicators, how do various employment indicators correlate, and what temporal trends and significant relationships can be identified in the context of regional employment disparities across country regions? To consider this, the following objectives will be met

The Objectives for the task are;

1. Conduct a thorough descriptive statistical analysis on employment indicators globally and across regions.
2. Evaluate correlations between chosen employment indicators to discern patterns influencing regional employment services.
3. Formulate and test two hypotheses to uncover significant relationships or differences within the dataset regarding disparities in employment across country regions.
4. Utilize regression analysis to examine relationships between indicators, justifying chosen techniques and validating them against existing literature.
5. Apply time series analysis to reveal temporal trends in employment indicators, explaining method choice and referencing relevant literature for validation.

2.0 Background Research

In the context of understanding global employment disparities, essential tasks involve using robust statistical methods like correlation analysis, hypothesis testing, and regression analysis. These methods help explore relationships, patterns, and influences on employment indicators across different regions.

Descriptive statistical analysis, including measures like Mean, Median, Mode, Standard Deviation, Skewness, and Kurtosis, forms the basis for understanding the distribution and central tendencies of employment indicators. Rodriguez and Chen (2015) extensively used these measures to portray the employment landscape in urbanized areas, showcasing the utility of descriptive statistics in characterizing employment patterns. Correlation analysis is vital for revealing connections between indicators. Previous research by Johnson et al. (2017) demonstrated its importance in uncovering relationships, such as the interplay between economic development and employment rates. This method identifies variables that move together, offering insights into potential causal relationships.

Formulating and testing hypotheses are crucial steps in examining employment disparities. Smith and Brown (2019) utilized hypothesis testing to analyze the impact of education on employment outcomes, revealing significant differences based on educational levels and providing valuable insights into education's role in shaping employment services. Regression analysis is a powerful tool for modelling relationships between variables. Kim and Patel (2018) applied regression techniques to examine the influence of GDP per capita on employment rates, illustrating a positive correlation and highlighting how economic development contributes to improved employment services. Time series analysis is crucial for examining temporal trends in employment indicators, allowing researchers to identify patterns, seasonality, and long-term trends, contributing to a comprehensive understanding of employment dynamics over time.

For a comprehensive analysis of global employment, Power BI's interactive dashboards enable effective communication and insightful exploration of complex data.

The design of interactive dashboards involves adherence to established principles of composition and layout. Tufte (2006) emphasizes the importance of maximizing the data-ink ratio, reducing non-essential elements, and employing efficient use of space for effective communication in visual displays. His work serves as a foundational reference for crafting dashboards that convey information concisely and coherently (Tufte, 2006).

The current perspectives on the design of single-screen dashboards underscore the significance of simplicity, clarity, and user-centric design. Few and Miller (2008) advocate for a minimalist approach to dashboard design, promoting the inclusion of only essential information to prevent cognitive overload. Their work contributes to understanding how design principles influence user engagement and comprehension in single-screen dashboards (Few & Miller, 2008). The methodology of designing and developing interactive dashboards demands a strategic approach. Steele et al. (2011) provide insights into the iterative process of dashboard development, emphasizing the need for collaboration between designers and end-users. Their methodology guides the effective integration of user feedback, ensuring that the interactive dashboard meets the audience's specific needs (Steele et al., 2011).

In the realm of Power BI, the works of Johnson (2016) delve into practical considerations and best practices for designing dashboards. His insights provide a valuable reference for leveraging Power BI's capabilities to create visually compelling and user-friendly interactive dashboards (Johnson, 2016).

2.1 Literature Review

This literature review provides valuable domain knowledge that informs the research objectives, allowing for a nuanced exploration of the factors influencing regional employment disparities. Scholars such as Becker and Huselid (2006) and Blinder (2009) have delved into the complexities of global employment dynamics, shedding light on the nuances of labour market trends, workforce participation, and the impact of globalization. Their work underscores the importance of a descriptive understanding of employment indicators, providing a foundation for the initial research objective. Studies by Rios-Salas and Trupp (2014) and Lee and Vivarelli (2006) offer insights into the correlations between socioeconomic factors and regional employment services. By exploring these connections, the research aims to discern patterns that influence employment disparities across diverse country regions.

Martin and Sunley (2015) and Glaeser et al. (2009), explored the drivers behind regional economic disparities. Their investigations into factors such as infrastructure development, educational systems, and innovation serve as a valuable backdrop for formulating and testing hypotheses related to employment variations. Building on the work of Rodriguez-Pose and Hardy (2015) and Fingleton et al. (2012), the research draws from literature exploring regional economic development strategies. This knowledge contributes to the design by providing insights into the regional policies and interventions that may impact employment trends across diverse geographical areas.

Temporal trends, as discussed by Autor (2019) and Blanchard and Wolfers (2000), provide a contextual understanding of how labour markets evolve. The application of time series analysis is informed by their insights, allowing for exploring temporal patterns in employment indicators. In synthesising these key insights from the literature, the research design is enriched with a solid foundation in the domain knowledge of employment dynamics, regional disparities, and the multifaceted factors shaping employment services.

Information regarding variations in employment service trends across different groups is straightforward to comprehend and is most effectively conveyed through a line chart, as exemplified by the visual representation provided by the World in Data Dashboard Visuals for this report. Leveraging colour tones for value distinction facilitates a more efficient search by users compared to other characteristics (Few, 2010). Bar charts are employed to prevent misinterpretation, enabling individuals with rudimentary graph-reading skills to promptly identify the highest value at the top and the lowest value at the bottom, as illustrated in this dashboard's visuals. Wexler et al. (2017)

explored dashboard design for monitoring economic indicators. Dashboards offer a visual representation of complex data, potentially revealing patterns and trends in employment services. Such visualization tools can serve as valuable instruments in elucidating the relationships between employment in services, wage and salary workers, and GDP per capita.

In addition to the academic discourse, online platforms and tools contribute to the exploration of employment data through interactive dashboards. While not exhaustive, sources like the World Bank's Databank (n.d.) and Microsoft's Power BI (Microsoft, 2022) provide interfaces for users to visualize and analyze employment indicators. These platforms leverage visualization techniques to transform raw data into actionable insights, fostering a dynamic understanding of employment services globally and across regions.

3.0 Preparation and Exploration of Data Set

Data was sourced from the World Bank (World Development Indicators, n.d.) comprises 15 countries in Sub-Saharan Africa, Europe, and East Asia (see Table 3). The selection spans 2002-2016. R is used for statistical analysis, including column deletion and renaming. Missing data columns are removed. Indicator definitions, from metadata (DataBank, n.d.), are detailed in the accompanying table. (For additional information, refer to the metadata zip folder.)

Region	**Sub-Saharan Africa**	**Europe**	**East Asia**
Countries	1. South Africa 2. Nigeria 3. Ghana 4. Kenya 5. Sierra Leone	1. Denmark 2. France 3. Germany 4. Italy 5. United Kingdom	1. China 2. Japan 3. Malaysia 4. Korea, Rep 5. Australia

Table 1 Countries and Regions

No	Indicators	Definition	Source
1	**Employment in services (% of total employment)**	Working-age individuals engaged in profit-driven activities, spanning wholesale/retail trade, transport, financing, and community services.	International Labour Organization, ILOSTAT database (September 2018).
2	**GDP per capita (constant 2010 US$)**	Gross domestic product per capita, calculated as GDP divided by midyear population, in constant 2010 U.S. dollars.	World Bank national accounts data, and OECD National Accounts data files.
3	**Labor force participation rate, total (%)**	Proportion of the population aged 15 and older economically active during a specified period, supplying labour for goods and services production.	International Labour Organization, ILOSTAT database (September 2018)
4	**Total employment, total (ages 15+)**	Total number employed ages 15 and over.	International Labour Organization, ILOSTAT database (March 2017)
5	**Unemployment, total (% of the total labor force)**	Share of the labour force without work but available for and seeking employment.	International Labour Organization, ILOSTAT database (September 2018)
6	**Unemployment, youth total (% of total labor force ages 15-24)**	Share of the labor force ages 15-24 without work but available for and seeking employment.	International Labour Organization, ILOSTAT database (September 2018).
7	**Urban population (% of total)**	People living in urban areas defined by national statistical offices.	United Nations Population Division. World Urbanization Prospects: 2018 Revision.
8	**Wage and salaried workers, total (% of total employment)**	Workers holding "paid employment jobs" with explicit or implicit employment contracts, not directly dependent on the unit's revenue.	International Labour Organization, ILOSTAT database (September 2018).

Table 2 Data Indicators Dictionary

The dataset intended for this study is loaded and examined using the Glimpse() function.

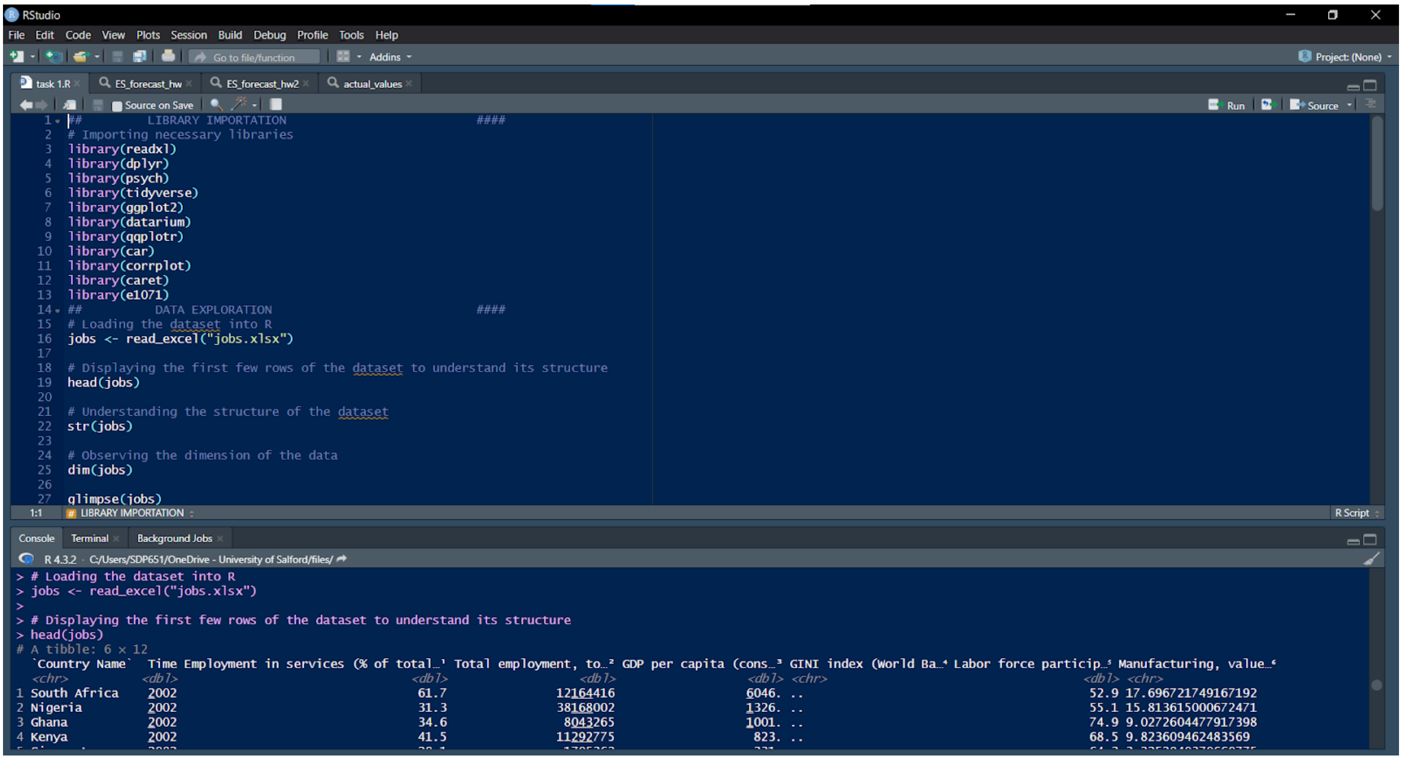


Figure 1. Loading of the Dataset

The dataset comprises 12 columns and 225 rows. For a preview of the dataset, please refer to Figure 2.

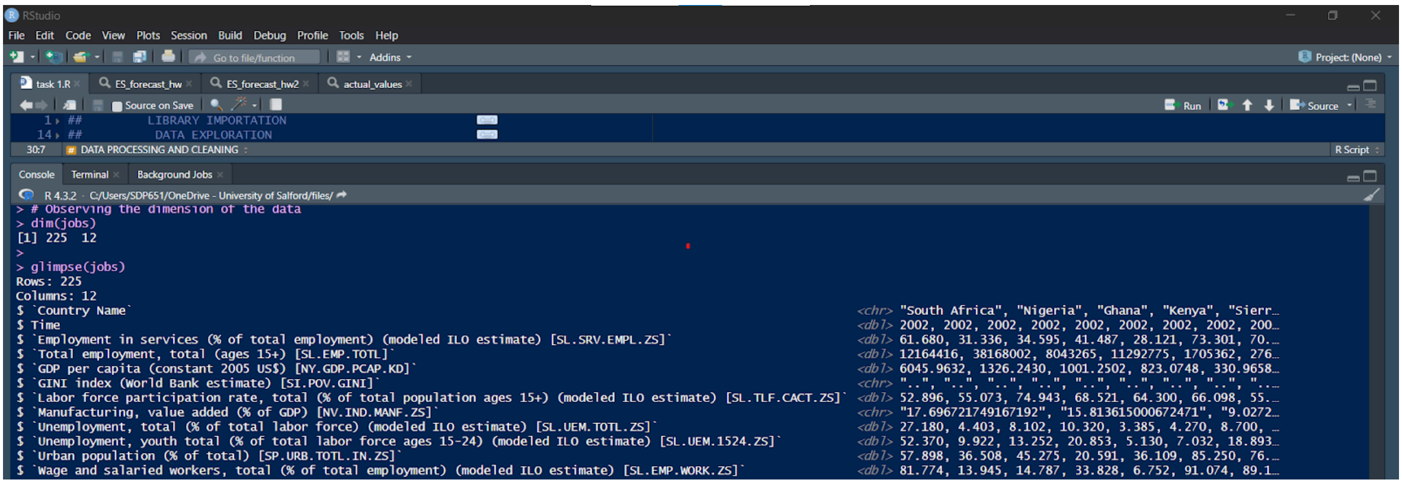


Figure 2. Glimpse of Dataset

```r
##      DATA PROCESSING AND CLEANING                    ####

# Renaming columns for easy encoding
jobs <- jobs %>%
  rename(
    'country' = 'Country Name',
    'year' = 'Time',
    'employment_services' = 'Employment in services (% of total employment) (modeled ILO estimate) [SL.SRV.EMPL.ZS]',
    'gdp_per_capita' = 'GDP per capita (constant 2005 US$) [NY.GDP.PCAP.KD]',
    'gini_index' = 'GINI index (World Bank estimate) [SI.POV.GINI]',
    'labor_force_participation_rate' = 'Labor force participation rate, total (% of total population ages 15+) (modeled ILO estimate) [SL.TLF.CACT.ZS]',
    'Manuf_Value' = 'Manufacturing, value added (% of GDP) [NV.IND.MANF.ZS]',
    'total_employment' = 'Total employment, total (ages 15+) [SL.EMP.TOTL]',
    'unemployment_total' = 'Unemployment, total (% of total labor force) (modeled ILO estimate) [SL.UEM.TOTL.ZS]',
    'unemployment_youth' = 'Unemployment, youth total (% of total labor force ages 15-24) (modeled ILO estimate) [SL.UEM.1524.ZS]',
    'urban_population' = 'Urban population (% of total) [SP.URB.TOTL.IN.ZS]',
    'wage_and_salaried_workers' = 'Wage and salaried workers, total (% of total employment) (modeled ILO estimate) [SL.EMP.WORK.ZS]'
  )
```

```
41:1   DATA EXPLORATION                                    R Script
Console   Terminal   Background Jobs
R 4.3.2 · C:/Users/SDP651/OneDrive - University of Salford/files/
>
> #Displaying Name change
> colnames(jobs)
 [1] "country"                "year"                           "employment_services"        "total_employment"        "gdp_per_capita"
 [6] "gini_index"             "labor_force_participation_rate" "Manuf_Value"                "unemployment_total"      "unemployment_youth"
[11] "urban_population"       "wage_and_salaried_workers"
>
> # Changing type from chr to dbl and checking the class of each column after the conversion
```

Figure 3. Changing column names for easy calling.

Excluding 'year' and 'country,' all columns are converted to numeric using the sapply function for consistent statistical analysis (R Programming Language, 2022). Details in Figures 3 and 4.

```r
# Changing type from chr to dbl and checking the class of each column after the conversion
cols.num <- 3:12
jobs[cols.num] <- sapply(jobs[cols.num], as.numeric)
sapply(jobs, class)

# Checking for missing data in each column using sapply
print("Columns with Missing Data:")
sapply(jobs, function(col) sum(is.na(col)))
```

Figure 4. Converting types

```
> sapply(jobs, class)
                country                           year        employment_services           total_employment             gdp_per_capita
            "character"                      "numeric"                  "numeric"                  "numeric"                  "numeric"
             gini_index labor_force_participation_rate                Manuf_Value         unemployment_total         unemployment_youth
              "numeric"                      "numeric"                  "numeric"                  "numeric"                  "numeric"
       urban_population      wage_and_salaried_workers
              "numeric"                      "numeric"
>
```

Figure 5. Converted types

Checking for missing data is vital. However, due to diverse country characteristics, imputing is impractical. Entries with missing data are removed.

```r
# Checking for missing data in each column using sapply
print("Columns with Missing Data:")
sapply(jobs, function(col) sum(is.na(col)))

# Dropping columns with missing values
jobs <- jobs %>%
  select(-which(colnames(jobs) %in% c(
    'Manuf_Value', 'gini_index'
  )))
```

```
427:1   MODEL2 ARIMA TIME SERIES                           R Script
Console   Terminal   Background Jobs
R 4.3.2 · C:/Users/SDP651/OneDrive - University of Salford/files/
[1] "Columns with Missing Data:"
> sapply(jobs, function(col) sum(is.na(col)))
                country                           year        employment_services           total_employment             gdp_per_capita
                      0                              0                          0                          0                          0
             gini_index labor_force_participation_rate                Manuf_Value         unemployment_total         unemployment_youth
                    139                              0                          2                          0                          0
       urban_population      wage_and_salaried_workers
                      0                              0
>
```

Figure 6. Missing Values

```r
# Assigning regions to countries
jobs$region[jobs$country %in% c("South Africa", "Nigeria", "Ghana", "Kenya", "Sierra Leone")] <- "Sub-Saharan Africa"
jobs$region[jobs$country %in% c("Denmark", "France", "Germany", "Italy", "United Kingdom")] <- "Europe"
jobs$region[jobs$country %in% c("China", "Japan", "Malaysia", "Korea, Rep.", "Australia")] <- "East Asia"
```

Figure 7. Countries grouped

in Africa exhibit the highest variance, followed by Asia, in the employment service sector's influence on employment. Similar patterns are observed in GDP per capita, urban population, wage and salaried workers, and total employment. Europe and East Asia display low variance and standard deviation, indicating favourable employment situations. Statistically evaluating this follows in the next section. A few outliers are noted, but their removal could compromise project objectives, so none are eliminated.

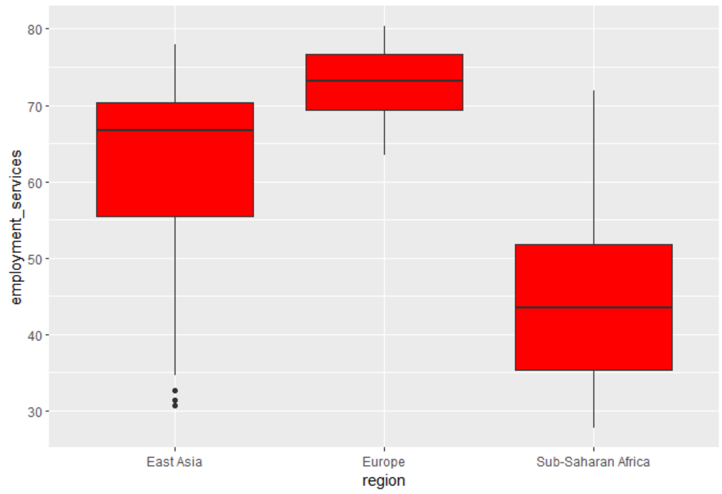
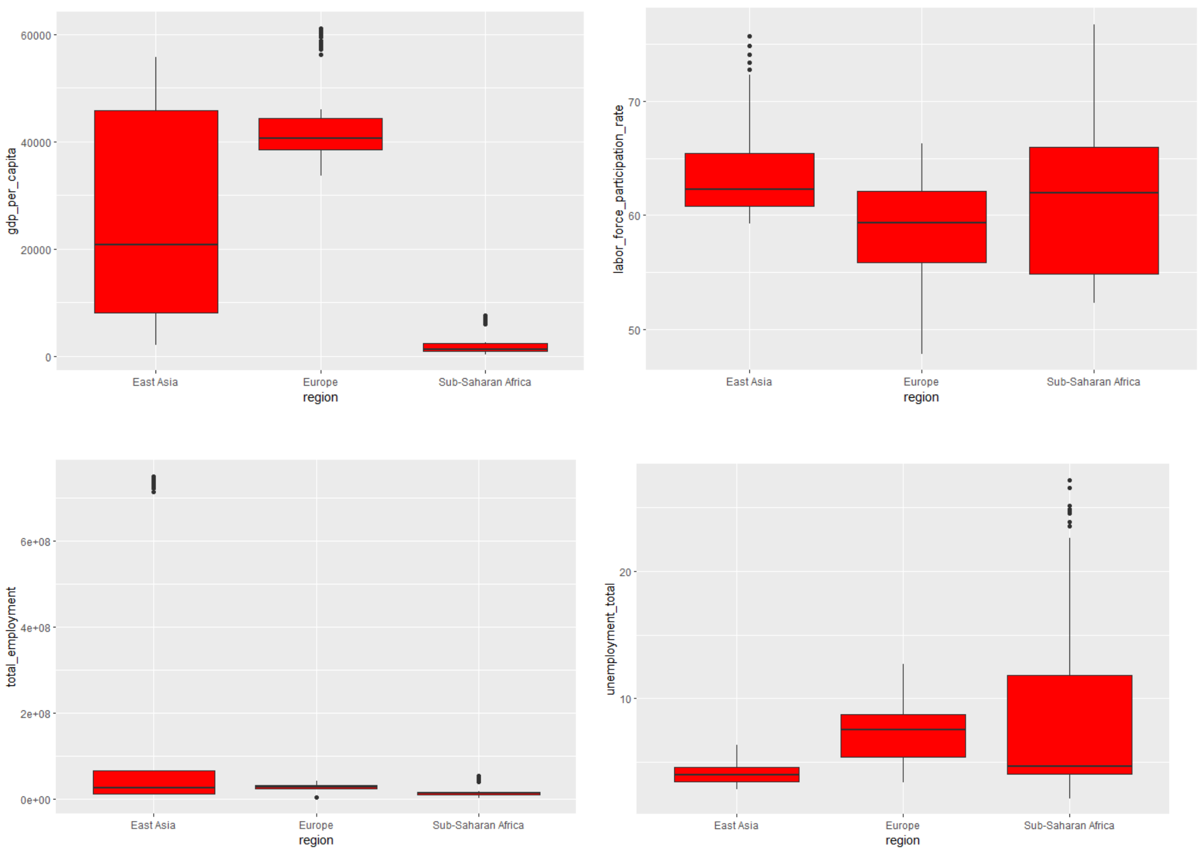


Figure 8. Employment services



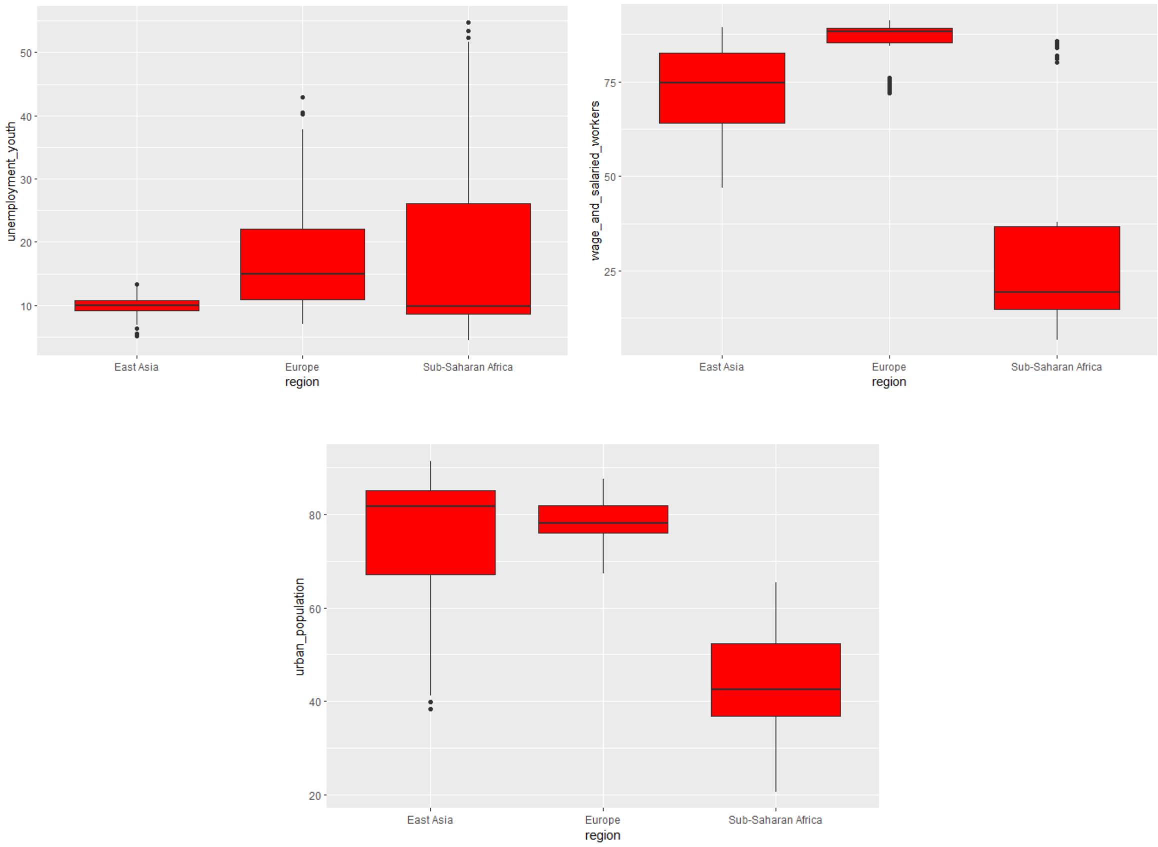


Figure 9. Other Indicators

Part Two: Statistical Analysis

1.0 Comprehensive Descriptive Statistical Analysis

Histograms (Figures 11 and 12) expose dataset shapes. Outliers, especially in the negatively skewed Employment Service distribution, hint at countries with high employment. Similar trends appear in GDP per capita, urban population, wage and salaried workers, and labour force participation rate distributions, with peaks leaning right.

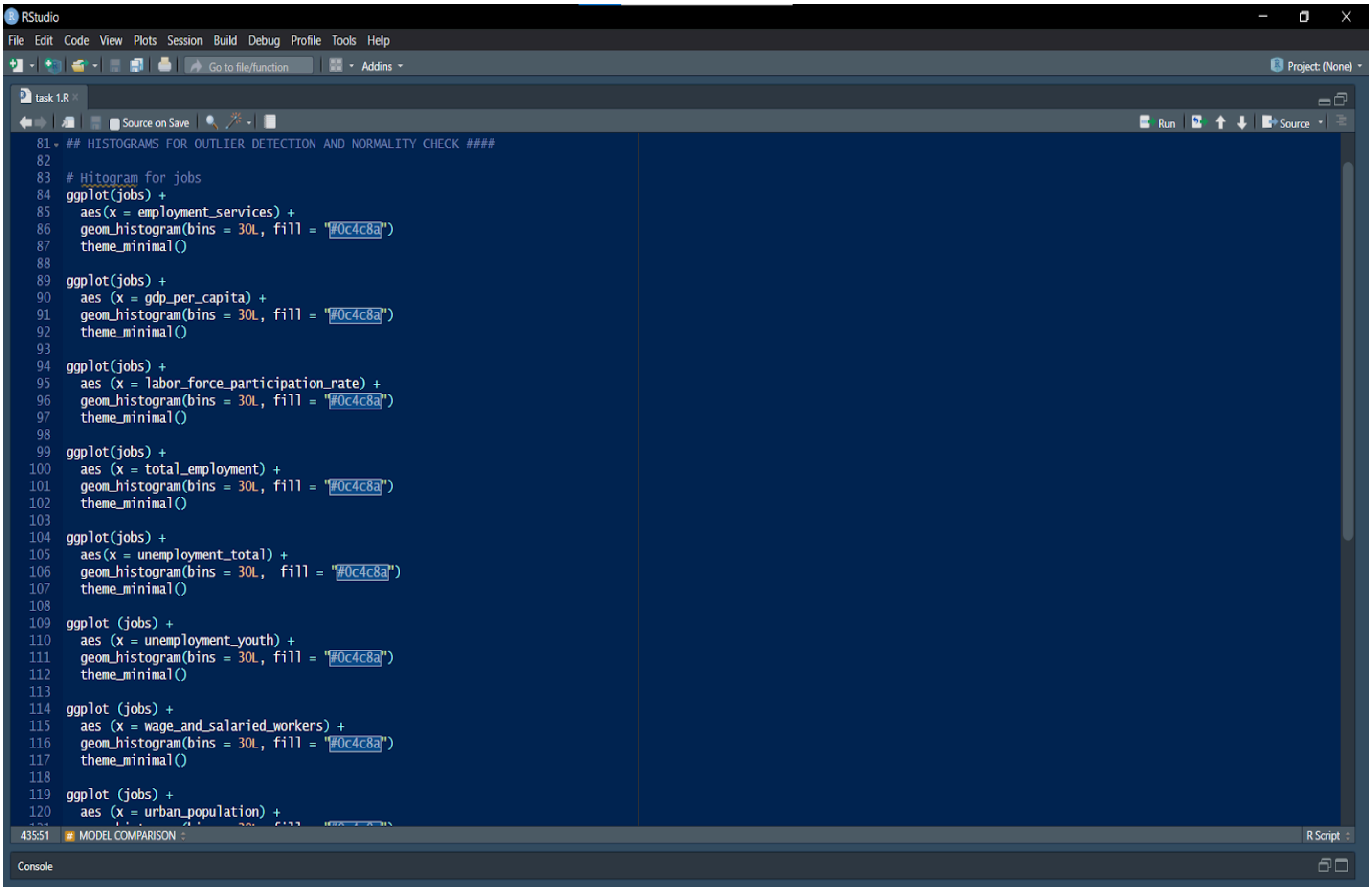


Figure 10. Histogram Code

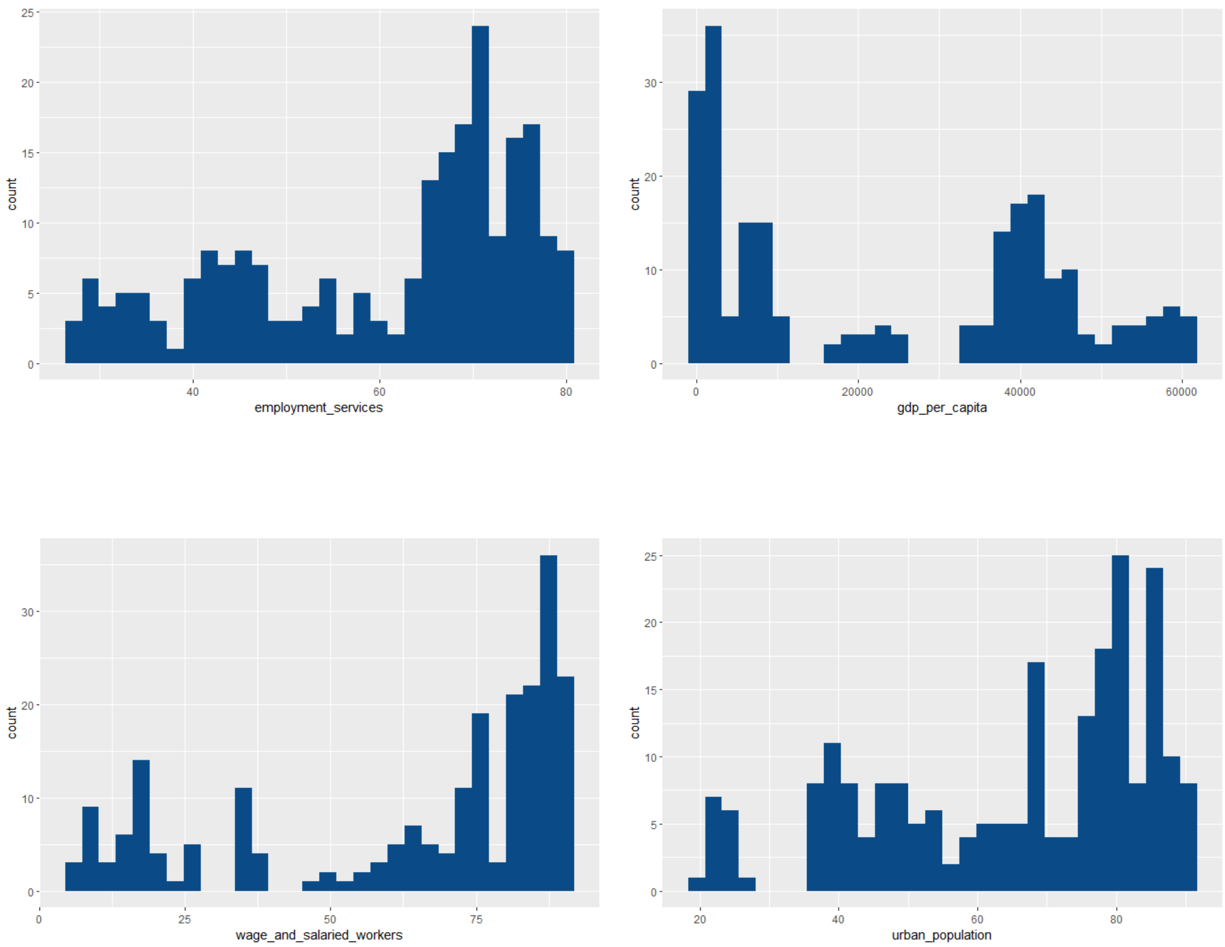
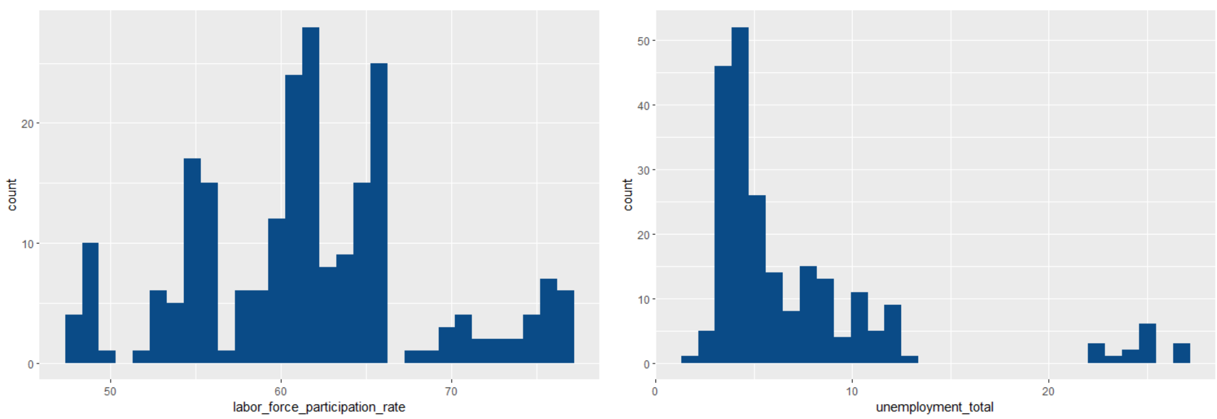


Figure 11. Histogram for the distribution of Employment Services, GDP per Capita, Wage and Salaried Workers and Urban Population



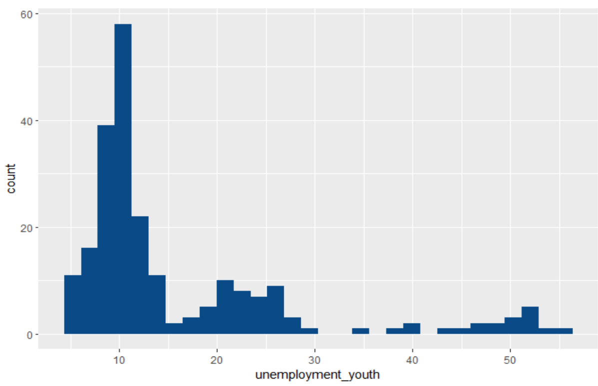


Figure 12. Histogram for the distribution of Labor Force Participation Rate, Unemployment Total and Unemployment Youth

a. Mean and Median

Here is the code for calculating the mean and median for each country.

```
# Summary statistics to explore all job indicators
summary_stats <- describe(jobs)
print(summary_stats)

# Mean of Indicators per country
each_country_mean <- jobs %>%
  group_by(country) %>%
  summarise(across(c(employment_services, gdp_per_capita,labor_force_participation_rate,
                    total_employment, unemployment_total, unemployment_youth,
                    urban_population, wage_and_salaried_workers), list(mean = mean)))

# Median of Indicators per country
each_country_median <- jobs %>%
  group_by(country) %>%
  summarise(across(c(employment_services, gdp_per_capita,labor_force_participation_rate,
                     total_employment, unemployment_total, unemployment_youth,
                     urban_population, wage_and_salaried_workers), list(median = median)))
```

Figure 13. Mean and Median Code Snippet

Tables 3 and 4 show that the country with the lowest mean and median value for Employment Services is Sierra Leone, with a mean of 30.575 and a median value of 29.642.

task 1.R | each_country_mean

Filter

	country	employment_services_mean	gdp_per_capita_mean	labor_force_participation_rate_mean	total_employment_mean	unemployment_total_mean	unemployment_youth_mean	urban_population_mean	wage_and_salaried_workers_mean
1	Sierra Leone	30.57527	411.1677	61.73993	2074391	4.086400	7.348333	38.51880	9.15020
2	Ghana	39.94453	1329.1363	75.97460	10513826	4.325400	8.219333	50.02567	20.40873
3	China	42.14673	4275.2518	71.87227	741349398	4.330333	9.860533	47.75753	57.26533
4	Kenya	45.23093	958.8756	65.93040	13486404	11.255467	24.171400	23.24440	35.79007
5	Nigeria	45.99793	2136.0835	54.96093	45816199	4.332067	9.851600	42.59433	17.40567
6	Malaysia	57.40760	8994.3562	61.85807	11910173	3.308000	10.726133	69.76153	74.92607
7	Italy	67.00753	35963.6055	48.61300	22414331	9.120000	29.704000	68.36653	74.25907
8	Korea, Rep.	67.39747	21206.8966	61.65867	24483378	3.470667	9.591400	81.49447	66.51027
9	South Africa	68.35360	7072.6735	53.59647	14357912	24.598667	50.203333	61.66113	83.41553
10	Japan	68.78753	44896.9645	60.33220	63740070	4.322000	7.992533	88.54760	85.76493
11	Germany	68.91800	41999.8281	59.29080	38679897	7.430000	10.164200	76.58020	88.46940
12	France	74.10880	40779.9619	55.99567	26861283	8.936000	21.456867	78.13187	88.64873
13	Denmark	75.87213	58841.2626	64.54720	2757989	5.689333	10.583000	86.55520	90.97760
14	Australia	75.96987	51339.7597	64.86567	10778359	5.379333	11.496667	85.04180	81.63680
15	United Kingdom	78.10487	39753.1139	61.89607	29799849	6.028667	15.715333	81.00560	86.07160

Table 3. Displays the mean values of the indicators per country, organized in ascending order based on the Employment Situation, from the lowest to the highest.

	country	employment_services_median	gdp_per_capita_median	labor_force_participation_rate_median	total_employment_median	unemployment_total_median	unemployment_youth_median	urban_population_median	wage_and_salaried_wo
1	Sierra Leone	29.642	387.811	61.956	2092194	4.140	7.384	38.466	
2	Ghana	40.904	1247.463	75.991	10391883	4.472	8.708	50.031	
3	China	42.056	4142.038	71.451	745981905	4.340	9.936	47.880	
4	Kenya	45.533	935.637	65.839	13160437	11.520	25.926	23.183	
5	Nigeria	48.697	2216.499	55.009	45594826	4.215	9.846	42.588	
6	Malaysia	58.050	8890.793	61.324	11482711	3.330	10.777	70.075	
7	Italy	67.071	35994.134	48.639	22406873	8.360	27.166	68.209	
8	Korea, Rep.	68.478	20843.135	61.860	24135565	3.530	9.718	81.634	
9	Germany	69.521	41831.867	59.358	38745038	7.520	9.831	76.771	
10	Japan	69.525	44995.494	60.456	63575830	4.350	8.090	89.989	
11	South Africa	69.677	7275.382	53.432	14491860	24.670	51.050	61.687	
12	France	74.400	41158.885	56.112	27117569	8.810	21.846	78.117	
13	Australia	75.625	51770.907	65.008	10855772	5.390	11.550	85.063	
14	Denmark	77.123	58788.082	65.323	2750509	6.010	10.798	86.654	
15	United Kingdom	79.278	39740.903	61.977	29597705	5.350	14.630	81.031	

Table 4. Displays the median values of the indicators per country, organized in ascending order based on the Employment Situation, from the lowest to the highest.

```r
# Mean of Indicators per region
each_region_mean <- jobs %>%
  group_by(region) %>%
  summarise(across(c(employment_services, gdp_per_capita,labor_force_participation_rate,
                   total_employment, unemployment_total, unemployment_youth,
                   urban_population, wage_and_salaried_workers), list(mean = mean)))

# Median of Indicators per region
each_region_median <- jobs %>%
  group_by(region) %>%
  summarise(across(c(employment_services, gdp_per_capita,labor_force_participation_rate,
                   total_employment, unemployment_total, unemployment_youth,
                   urban_population, wage_and_salaried_workers), list(median = median)))
```

Figure 14. Exhibition of the code for categorising regions based on indicators.

East Asia has the highest mean and median values for Employment Services and Urban Population, while Europe leads in GDP per Capita and Wage and Salaried Workers. Sub-Saharan Africa consistently has the lowest values, indicating potential disparities needing attention and interventions.

region	employment_services_mean	gdp_per_capita_mean	labor_force_participation_rate_mean	total_employment_mean	unemployment_total_mean	unemployment_youth_mean	urban_population_mean	wage_and_salaried_workers_mean
East Asia	62.34184	26142.646	64.11737	170452276	4.162067	9.933453	74.52059	73.22068
Europe	72.80227	43467.554	58.06855	24102670	7.440800	17.524680	78.12788	85.68528
Sub-Saharan Africa	46.02045	2381.587	62.44047	17249746	9.719600	19.958800	43.20887	33.23404

Table 5. Displays the mean values of the indicators per region

	region	employment_services_median	gdp_per_capita_median	labor_force_participation_rate_median	total_employment_median	unemployment_total_median	unemployment_youth_median	urban_population_median	wage_and_salaried_
1	East Asia	66.715	20843.135	62.284	24135565	4.030	9.958	81.634	
2	Europe	73.221	40638.334	59.358	27117569	7.540	15.020	78.117	
3	Sub-Saharan Africa	43.506	1247.463	61.956	12915103	4.703	9.905	42.588	

Table 6. Displays the median values of the indicators per region

b. Standard Deviation

The analysis of employment services across regions utilizes standard deviation to assess data spread. Higher standard deviation suggests greater variability, highlighting potential disparities in employment services within regions.

```r
# Standard deviation of Indicators per region
each_region_sd <- jobs %>%
  group_by(region) %>%
  summarise(across(c(employment_services, gdp_per_capita, labor_force_participation_rate,
                     total_employment, unemployment_total, unemployment_youth,
                     urban_population, wage_and_salaried_workers), list(sd = sd)))

# Standard deviation of Indicators per region
each_country_sd <- jobs %>%
  group_by(country) %>%
  summarise(across(c(employment_services, gdp_per_capita, labor_force_participation_rate,
                     total_employment, unemployment_total, unemployment_youth,
                     urban_population, wage_and_salaried_workers), list(sd = sd)))
```

Figure 15. Code for summarizing standard deviation per region and country based on indicators.

	region	employment_services_sd	gdp_per_capita_sd	labor_force_participation_rate_sd	total_employment_sd	unemployment_total_sd	unemployment_youth_sd	urban_population_sd	wage_and_salaried_workers_sd
1	East Asia	12.412628	19132.300	4.341213	288065895	0.8663975	1.556434	15.263683	10.79420
2	Europe	4.691684	8177.877	5.613917	12039719	2.2063604	8.500009	6.058834	6.00518
3	Sub-Saharan Africa	13.194614	2448.256	8.291679	15240984	8.0539686	16.562738	13.060998	26.82426

Table 7. Standard Deviation of Indicators per Region

From Table 7 above, Sub-Saharan Africa exhibits the highest standard deviation in Employment Services, suggesting significant disparities. Europe shows moderate variability in GDP per Capita and Wage and Salaried Workers. East Asia presents considerable variability in Urban Population.

	country	employment_services_sd	gdp_per_capita_sd	labor_force_participation_rate_sd	total_employment_sd	unemployment_total_sd	unemployment_youth_sd	urban_population_sd	wage_and_salaried_workers_sd
1	Australia	1.048045	3083.42620	0.6933770	859828.48	0.6208918	1.3240344	0.5094294	1.1632233
2	China	8.020975	1599.35855	1.9326736	11495691.52	0.2402406	0.7164029	5.8637536	5.5943810
3	Denmark	2.372625	1604.76021	1.6007583	54680.06	1.4171373	2.6690444	0.7951669	0.1740426
4	France	2.009105	907.45666	0.2959234	572225.55	0.9420631	2.5006414	1.1257305	0.3753185
5	Germany	1.922065	2729.44238	0.9804465	1630844.80	2.3569501	2.5725219	0.6581841	0.5428721
6	Ghana	3.529155	262.56111	0.6507862	1570757.68	1.7836583	2.7584218	3.0330815	4.7651796
7	Italy	2.189913	1551.23010	0.4400532	399754.10	2.1647038	7.6464069	0.7625578	1.2063375
8	Japan	1.903715	1618.78754	0.6176224	562842.77	0.6930182	1.5476073	3.2975381	2.3539842
9	Kenya	1.978364	103.10478	0.9116931	1756557.65	0.6766228	2.2003616	1.7539201	1.2862995
10	Korea, Rep.	2.370335	2839.68743	0.6258826	1274011.76	0.2192020	0.6251318	0.4819788	2.8684445
11	Malaysia	2.400435	1207.58482	1.5149748	1646728.07	0.2389920	0.5419270	3.5362288	0.9147662
12	Nigeria	7.388268	386.77695	0.1416901	5204915.04	0.7992491	1.0485007	3.9119702	1.6427321
13	Sierra Leone	2.810126	71.49878	2.7575911	178138.56	0.4979640	1.7086818	1.6833093	1.3513033
14	South Africa	3.341856	538.49768	0.9744110	1144763.54	1.5289952	2.6368456	2.3787355	1.7935979
15	United Kingdom	1.999474	1359.10573	0.3386939	1115518.59	1.3244668	3.8406489	1.2246793	0.9622009

Table 8. Standard Deviation of Indicators per country

China exhibits substantial diversity in Employment Services and GDP per Capita, while Australia showcases minimal variability in Urban Population and Wage and Salaried Workers.

c. Skewness and Kurtosis

Skewness gauges the asymmetry of a dataset, with values between -0.5 and 0.5 suggesting symmetry. Moderate skewness falls between -1 to -0.5 or 0.5 to 1, while highly skewed data has values less than -1 or greater than 1. Kurtosis evaluates tail behaviour: mesokurtic (normal) exhibits a kurtosis near 0, leptokurtic (heavy-tailed) has a positive kurtosis, and platykurtic (light-tailed) displays a negative kurtosis (Skewness and Kurtosis in Statistics, n.d.).

```r
# Skewness of Indicators per country
each_country_skew <- jobs %>%
  group_by(country) %>%
  summarise(across(c(employment_services, gdp_per_capita, labor_force_participation_rate,
                     total_employment, unemployment_total, unemployment_youth,
                     urban_population, wage_and_salaried_workers), list(skew = skew)))

# Kurtosis of indicators for per country
each_country_kurtosis <- jobs %>%
  group_by(country) %>%
  summarise(across(c(employment_services, gdp_per_capita, labor_force_participation_rate,
                     total_employment, unemployment_total, unemployment_youth,
                     urban_population, wage_and_salaried_workers), list(kurtosis = kurtosis)))

# Skewness of Indicators per region
each_region_skew <- jobs %>%
  group_by(region) %>%
  summarise(across(c(employment_services, gdp_per_capita, labor_force_participation_rate,
                     total_employment, unemployment_total, unemployment_youth,
                     urban_population, wage_and_salaried_workers), list(skew = skew)))

# Kurtosis of indicators per region
each_region_kurtosis <- jobs %>%
  group_by(region) %>%
  summarise(across(c(employment_services, gdp_per_capita, labor_force_participation_rate,
                     total_employment, unemployment_total, unemployment_youth,
153:13  DESCRIPTIVE STATISTICAL ANALYSIS
```

Figure 16. Code for extracting the Skewness and Kurtosos per region and country based on indicators.

Table 8 illustrates skewness in regional employment services. East Asia has negative skewness, Europe shows symmetry, Sub-Saharan Africa indicates positive skewness, and Australia suggests upward trends. China's distribution is moderately skewed, Denmark is negatively skewed, and France, Germany, Ghana, and Italy show diverse skewness patterns.

	region	employment_services_skew	gdp_per_capita_skew	labor_force_participation_rate_skew	total_employment_skew	unemployment_total_skew	unemployment_youth_skew	urban_population_skew	wage_and_salaried_workers_skew
1	East Asia	-0.9349343	0.212281	1.1066806	1.4549256	0.6402849	-0.5792958	-0.9764773	-0.5083297
2	Europe	-0.1524360	1.147862	-0.5439373	-0.7214238	0.2746024	0.9982291	-0.2718319	-1.2360518
3	Sub-Saharan Africa	0.6176605	1.291971	0.4898859	1.2156414	1.0901766	1.0274326	-0.1102668	1.0958209

	country	employment_services_skew	gdp_per_capita_skew	labor_force_participation_rate_skew	total_employment_skew	unemployment_total_skew	unemployment_youth_skew	urban_population_skew	wage_and_salaried_workers_skew
1	Australia	0.65290224	-0.33234079	-0.73161306	-0.26883710	-0.26689746	-0.51129041	-0.087003045	-0.29935394
2	China	0.06109496	0.14416583	0.52203521	-1.10665296	-0.70739548	-0.22417228	-0.044702399	-0.47744318
3	Denmark	-0.10436611	-0.14227788	-0.41507437	0.44278959	-0.13955049	0.03104387	-0.245259227	0.05524573
4	France	-0.17037138	-0.61942068	-0.78772034	-1.08267210	-0.09997128	-0.22053162	0.023495791	0.31535604
5	Germany	-0.51906042	-0.03211511	-0.37552146	0.02596178	0.15280355	0.51879255	-0.504096449	0.28712527
6	Ghana	-0.48861522	0.24115342	-0.14249710	0.04947471	0.43861780	0.00928367	-0.005554782	0.13089049
7	Italy	-0.16474193	-0.14717596	0.20656271	-0.42140530	0.34269594	0.45734841	0.531983581	-0.43455939
8	Japan	-0.42644236	-0.03647675	0.07169912	0.18809827	-0.07612387	-0.35987905	-0.791226837	0.17757390
9	Kenya	-0.47000663	0.27748693	1.54083711	0.32789957	-0.07530901	-0.27384011	0.080409575	0.26347706
10	Korea, Rep.	-0.46040988	-0.09719182	-0.12053804	0.37197889	-0.28840368	-0.36969905	-1.266942062	-0.17747406
11	Malaysia	-0.29571475	0.13440657	0.63053912	0.32415885	-0.10938302	-0.14458361	-0.170531408	0.75053257
12	Nigeria	-0.59856601	-0.76177295	-0.32271536	0.05561443	2.55713333	1.43956562	0.001210703	-0.60816668
13	Sierra Leone	0.77192637	0.91114063	-0.08078314	-0.44184231	-0.17467685	-0.06148019	0.094030481	-0.30983360
14	South Africa	-0.50814573	-0.76527647	0.55779762	-0.43953581	0.17942776	-0.21110545	-0.023907808	-0.37248433
15	United Kingdom	-0.26452433	-0.12556086	-0.58393086	0.54568213	0.45938205	0.23624030	-0.044290150	-0.18090086

Table 8. Skewness of Indicators per region and per country

	region	employment_services_kurtosis	gdp_per_capita_kurtosis	labor_force_participation_rate_kurtosis	total_employment_kurtosis	unemployment_total_kurtosis	unemployment_youth_kurtosis	urban_population_kurtosis	wage_and_salari
1	East Asia	0.01763411	-1.669869416	0.03335406	0.1463739	-0.5199419	1.0123562	-0.3455168	
2	Europe	-1.10711167	-0.146690392	-0.85077020	-0.5509707	-0.6682036	0.6871070	-0.7992534	
3	Sub-Saharan Africa	-0.65272279	0.001187071	-1.13969198	0.1433314	-0.3724511	-0.5107414	-0.9494100	

	country	employment_services_kurtosis	gdp_per_capita_kurtosis	labor_force_participation_rate_kurtosis	total_employment_kurtosis	unemployment_total_kurtosis	unemployment_youth_kurtosis	urban_population_kurtosis	wage_and_salari
1	Australia	-1.0108246	-1.2264445	-0.9236340	-1.3721969	-1.0200433	-0.8208189	-1.4682462	
2	China	-1.4659680	-1.4957170	-1.0740815	-0.1805708	-0.1493153	-1.0582766	-1.4467840	
3	Denmark	-1.9930101	-1.2770568	-1.5902381	-1.2423868	-1.4913973	-1.7460773	-1.4245824	
4	France	-1.3160201	-0.9902427	-0.8038943	-0.1398981	-0.8463506	-1.5562095	-1.4318710	
5	Germany	-1.1655861	-1.6650322	-1.3574760	-1.3332012	-1.5358535	-0.9681751	-1.3859300	
6	Ghana	-1.3495404	-1.7088444	-1.6085887	-1.3996268	-0.7376578	-1.2008238	-1.4468618	
7	Italy	-1.4329729	-1.5542987	-0.4010165	-0.4245026	-1.4434656	-1.4300824	-1.0559845	
8	Japan	-1.3608388	-1.1108072	-0.8789559	-1.0229998	-1.2719883	-1.1422502	-0.8964466	
9	Kenya	-0.9576470	-1.3201284	1.9304560	-1.3994086	-1.7897448	-1.8708177	-1.4228619	
10	Korea, Rep.	-1.4193371	-1.4430342	-1.4134384	-1.3108222	-1.6178874	-1.0115871	0.3594915	
11	Malaysia	-1.2165834	-1.2758120	-1.3173101	-1.4609816	-1.2562571	-1.1560640	-1.4390908	
12	Nigeria	-1.0114595	-0.6319395	-1.5571628	-1.5461083	6.1117286	3.6347874	-1.4563972	
13	Sierra Leone	-0.9165025	-0.4064426	-1.6548404	-0.6255611	-1.6620971	-1.7133839	-1.4919253	
14	South Africa	-1.3019780	-1.0462680	-0.6112382	-0.8525030	-0.9880359	-1.0510958	-1.4420329	
15	United Kingdom	-1.7156465	-0.9605977	-1.0063201	-0.6512786	-1.6679517	-1.5956895	-1.4356855	

Table 9. Kurtosis of Indicators per region and per indicators

Table 9 depicts regional employment services distributions. Europe's symmetrical distribution signals stability, while Sub-Saharan Africa shows a balanced pattern. East Asia displays mixed tail characteristics.

2.0 Correlation Analysis

Unlike Pearson, the Spearman Ranks method, using ranks, is robust to outliers and skewed distributions, especially beneficial with non-normal data.(Joost et al., 2016). Given the presence of outliers and non-normal distribution (refer to the next section), Spearman ranks are utilized.

```r
##   CORRELATION ANALYSIS FOR THE INDICATORS          ####

# Removing character value 'country' and 'region' to find correlation
jobs_new <- jobs[, -which(names(jobs) %in% c('country', 'region'))]
glimpse(jobs_new)

# Calculate Spearman's rank correlation for all variables Correlation plot
cor(jobs_new, method = 'spearman')

# Correlation plot
corrplot(cor(jobs_new),
         type = "lower",
         tl.col = "black",
         method = 'number',
         tl.srt = 45)

# Correlation tests with 'employment_services'
cor.test(jobs_new$employment_services, jobs_new$gdp_per_capita, method = 'spearman')
cor.test(jobs_new$employment_services, jobs_new$unemployment_youth, method = 'spearman')
cor.test(jobs_new$employment_services, jobs_new$total_employment, method = 'spearman')
cor.test(jobs_new$employment_services, jobs_new$unemployment_total, method = 'spearman')
cor.test(jobs_new$employment_services, jobs_new$urban_population, method = 'spearman')
cor.test(jobs_new$employment_services, jobs_new$labor_force_participation_rate, method = 'spearman')
cor.test(jobs_new$employment_services, jobs_new$wage_and_salaried_workers, method = 'spearman')
```

Figure 17. Code for processing categorical columns and correlation analysis.

The code in Figure 17 processes a data frame, excluding "country." It calculates Spearman's rank correlation using cor(data), creating a correlation plot with corrplot. The visual shows correlations in a lower triangular form with numeric values. It performs Spearman's rank correlation tests on 'employment_services' and indicators in 'jobs_new,' revealing relationships with all variables. The method parameter ensures adherence to Spearman's correlation method.

```
        Spearman's rank correlation rho

data:  jobs_new$employment_services and jobs_new$gdp_per_capita
S = 227799, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.880005
```

```
        Spearman's rank correlation rho

data:  jobs_new$employment_services and jobs_new$urban_population
S = 305224, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8392206
```

```
        Spearman's rank correlation rho

data:  jobs_new$employment_services and jobs_new$wage_and_salaried_workers
S = 261564, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8622189
```

```
        Spearman's rank correlation rho

data:  jobs_new$employment_services and jobs_new$labor_force_participation_rate
S = 2303061, p-value = 0.001297
alternative hypothesis: true rho is not equal to 0
sample estimates:
       rho
-0.2131591
```

Figure 18. Correlation test

Figure 19. Correlation plot

Spearman's rank correlation matrix reveals strong positive correlations of employment services with the urban population, GDP per capita, and wage and salaried workers, while labour force participation has a weak negative correlation. A weak negative correlation with labour force participation could worsen regional disparities, leading to challenges such as a limited workforce pool for employment opportunities in certain countries and regions compared to others.

3.0 Hypothesis Testing

Figure 20 illustrates density plots of mean employment services in East Asia, Europe, and Sub-Saharan Africa. Overlapping plots prompt a hypothesis test to assess the statistical significance of regional differences in employment services.

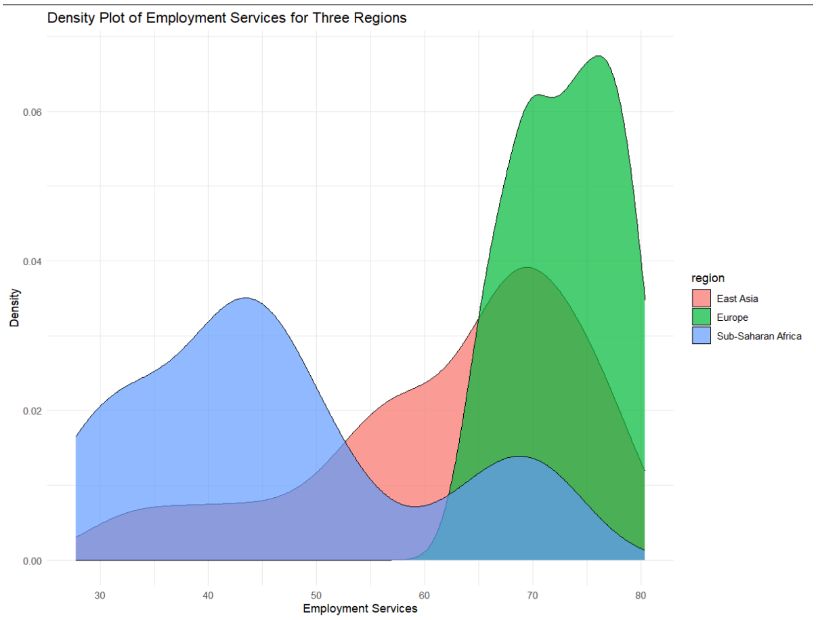


Figure 20. Employment Services for Tree Regions

Hypothesis

Null Hypothesis (Ho): There is no notable disparity in the mean levels of employment services across distinct regions (uEurope = uAsia = uSub-Saharan Africa).

Alternative Hypothesis (Hi): At least one region among the three exhibits a distinct mean in employment services.

Confidence Level (Alpha): 0.05.

Normality Check

To assess mean employment across multiple regions, a hypothesis test for comparing groups is crucial. Tests like ANOVA assume data normality, checked beforehand. Figure 21's Quantile-Quantile plot indicates a non-normal distribution due to substantial employment variations across regions over time.

Europe's mean life expectancy points closely align with and resemble a straight line in Figure 22, suggesting an approximate normal distribution. Conversely, in Asia and Africa, few points adhere to the line, signifying non-normal distribution in figures 23 and 24. To validate this, a Shapiro-Wilk test for normality is performed.

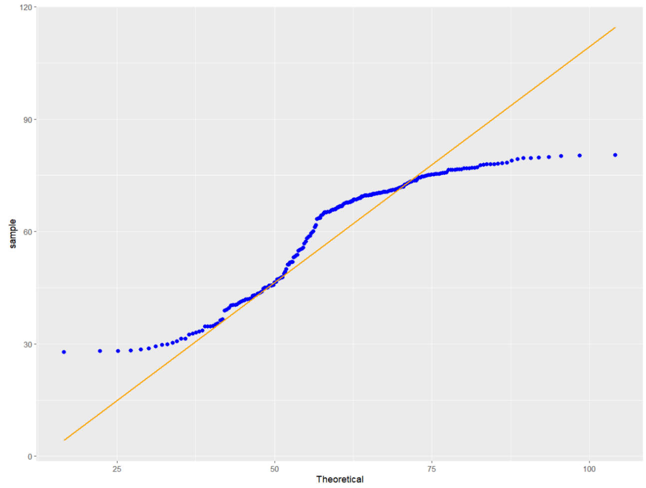


Figure 21. "Q-Q Plot Across the Three Regions

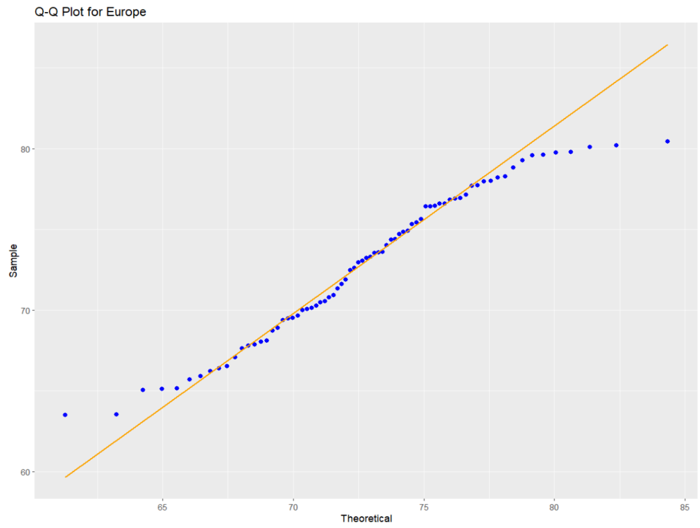


Figure 22. Q-Q Plot for Europe

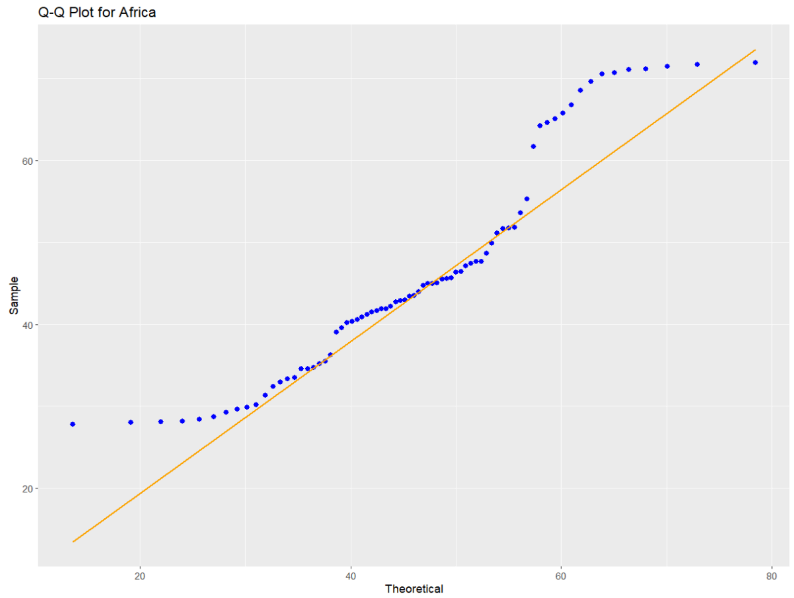


Figure 23. Q-Q Plot for Africa

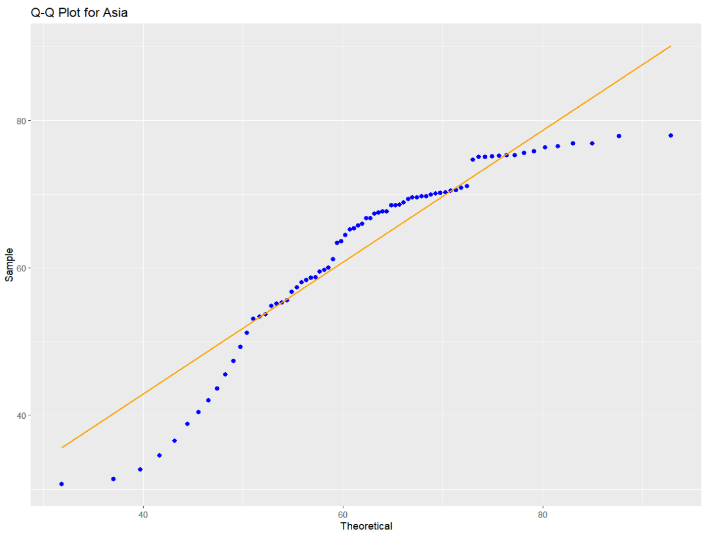


Figure 24. Q-Q Plot for Asia

The Shapiro-Wilk test examines the normal distribution of data, assuming a null hypothesis that the data is not normal and an alternative hypothesis that it is. A p-value above 0.05 suggests normal distribution. In the Employment Service mean normality test results, All region p-values indicate not normally distributed.

```
> shapiro_test_results
jobs$region: East Asia

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.90049, p-value = 2.258e-05

--------------------------------------------------------------------
jobs$region: Europe

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.96251, p-value = 0.02569

--------------------------------------------------------------------
jobs$region: Sub-Saharan Africa

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.91068, p-value = 6.067e-05
```

Figure 25. Test result

Normality Transformation

Ensuring accurate results in parametric hypothesis testing requires meeting the normality assumption. Three transformation methods will be tested for this purpose:

1. Log Transformation

2. Square Root

3. Cube Root

```
> # Shapiro-Wilk Test on Log-transformed data
> by(jobs$log_employment_services, jobs$region, shapiro.test)
jobs$region: East Asia

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.8397, p-value = 1.526e-07

---------------------------------------------------------------------
jobs$region: Europe

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.96024, p-value = 0.01892

---------------------------------------------------------------------
jobs$region: Sub-Saharan Africa

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.94673, p-value = 0.003313
```

```
> # Shapiro-Wilk Test on Square Root-transformed data
> by(jobs$sqrt_employment_services, jobs$region, shapiro.test)
jobs$region: East Asia

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.87303, p-value = 2.001e-06

---------------------------------------------------------------------
jobs$region: Europe

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.96156, p-value = 0.02259

---------------------------------------------------------------------
jobs$region: Sub-Saharan Africa

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.93296, p-value = 0.0006476
```

```
> # Shapiro-Wilk Test on Cube Root-transformed data
> by(jobs$cbrt_employment_services, jobs$region, shapiro.test)
jobs$region: East Asia

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.86255, p-value = 8.573e-07

---------------------------------------------------------------------
jobs$region: Europe

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.96116, p-value = 0.02141

---------------------------------------------------------------------
jobs$region: Sub-Saharan Africa

        Shapiro-Wilk normality test

data:  dd[x, ]
W = 0.93858, p-value = 0.00124

>
```

Figure 26. Transformation Results

In all cases, the p-values are very low, suggesting that even after transformations, the data for the transformed variables does not follow a normal distribution. As the data resisted normalization through various transformations, a non-parametric hypothesis testing method will be employed.

HYPOTHESIS 1: Kruskal-Wallis Test

The Kruskal-Wallis test, a non-parametric ANOVA, examines means between groups with no requirement for normal distribution. It is robust to variance assumptions. To proceed, confirmation of the following assumptions is necessary:

1. All samples are independent: Each observation is randomly selected, ensuring independence.

2. More than 2 categorical groups: (Europe, East Asia, Sub-Saharan Africa)

3. Groups are factors: As seen in the exploratory data analysis, groups are categorical factors.

4. Dependent variable is continuous: Average life expectancy is a continuous variable.

All assumptions are satisfied.

```
> # Kruskal-Wallis Test
> kruskal_test_result <- kruskal.test(employment_services ~ region, data = jobs)
>
> # Display the Kruskal-Wallis Test result
> print(kruskal_test_result)

        Kruskal-Wallis rank sum test

data:  employment_services by region
Kruskal-Wallis chi-squared = 109.45, df = 2, p-value < 2.2e-16
```

Figure 27. Kruskal-Wallis Test Result

```
> # Pairwise Wilcoxon Tests with Bonferroni correction
> pairwise_wilcox_test_result <- pairwise.wilcox.test(
+   jobs$employment_services, jobs$region,
+   p.adjust.method = "bonferroni"
+ )
>
> # Display the Pairwise Wilcoxon Tests result
> print(pairwise_wilcox_test_result)

        Pairwise comparisons using Wilcoxon rank sum test with continuity correction

data:  jobs$employment_services and jobs$region

                   East Asia Europe
Europe             1.3e-08   -
Sub-Saharan Africa 5.5e-10   < 2e-16

P value adjustment method: bonferroni
>
```

Figure 28. Wilcoson Result

The Kruskal-Wallis test, assessing mean ranks of employment services across regions, resulted in a chi-squared value of 109.45 with 2 degrees of freedom. The p-value, less than 2.2e-16, strongly rejects the null hypothesis, indicating significant differences in employment services distribution among the regions.

Pairwise Wilcoxon tests with Bonferroni correction were employed to discern specific group differences in employment services. Significant disparities were found in all region pairs: East Asia vs. Europe (p = 1.3e-08), East Asia vs. Sub-Saharan Africa (p = 5.5e-10), and Europe vs. Sub-Saharan Africa (p < 2e-16). Bonferroni adjustment was applied to control the overall Type I error rate, affirming the statistical significance of employment services differences between each region pair.

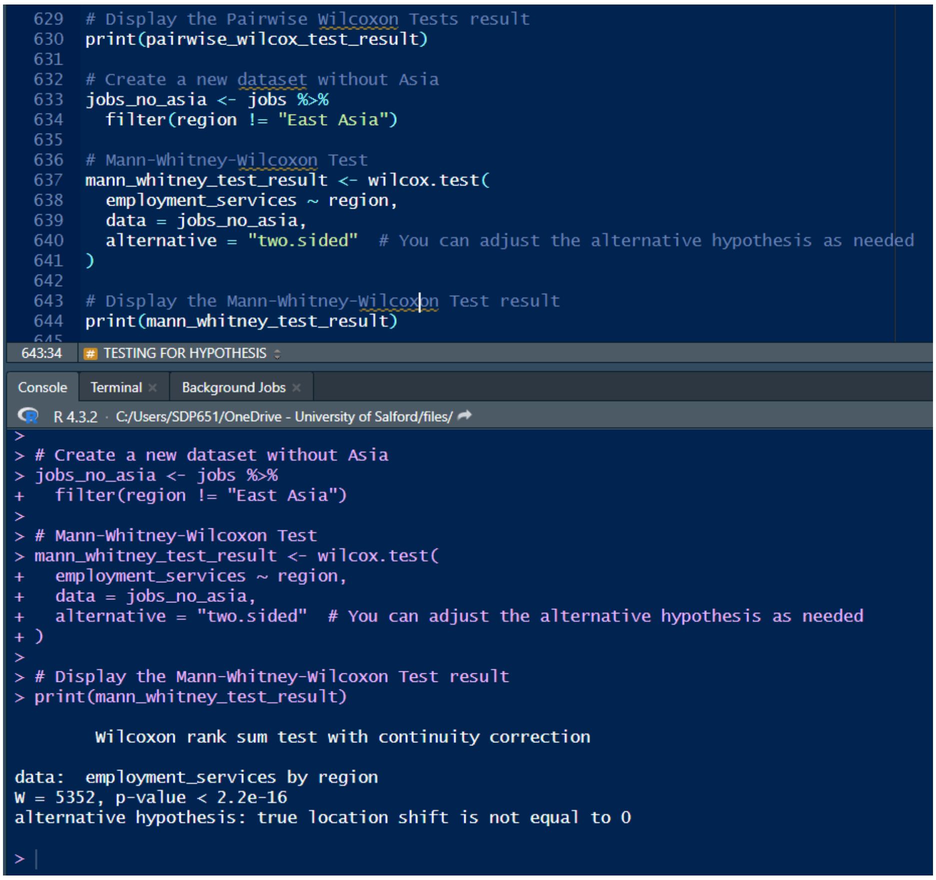
HYPOTHESIS 2: Mann-Whitney-Wilcoxon Test

Assumptions: Independent samples

Null Hypothesis (Ho): There is no significant difference in the means of employment services between Sub-sahran Africa and Europe (uEurope = uSub-Saharan Africa).

Alternative Hypothesis (Hi): The mean employment services in these two regions are significantly different.

Despite Figure 20's density plots showing apparent differences in the means of employment services between Europe and Sub-Saharan Africa, there is substantial overlap in values. Given this, Utilizing the Mann-Whitney-Wilcoxon Test—a non-parametric alternative to the parametric Student t-test, Hypothesis 2 aims to assess whether the mean employment services between Sub-Saharan Africa and Europe significantly differ.



```
# Display the Pairwise Wilcoxon Tests result
print(pairwise_wilcox_test_result)

# Create a new dataset without Asia
jobs_no_asia <- jobs %>%
  filter(region != "East Asia")

# Mann-Whitney-Wilcoxon Test
mann_whitney_test_result <- wilcox.test(
  employment_services ~ region,
  data = jobs_no_asia,
  alternative = "two.sided"  # You can adjust the alternative hypothesis as needed
)

# Display the Mann-Whitney-Wilcoxon Test result
print(mann_whitney_test_result)
```

```
643:34   TESTING FOR HYPOTHESIS
Console  Terminal  Background Jobs
R 4.3.2 · C:/Users/SDP651/OneDrive - University of Salford/files/
>
> # Create a new dataset without Asia
> jobs_no_asia <- jobs %>%
+   filter(region != "East Asia")
>
> # Mann-Whitney-Wilcoxon Test
> mann_whitney_test_result <- wilcox.test(
+   employment_services ~ region,
+   data = jobs_no_asia,
+   alternative = "two.sided"  # You can adjust the alternative hypothesis as needed
+ )
>
> # Display the Mann-Whitney-Wilcoxon Test result
> print(mann_whitney_test_result)

        Wilcoxon rank sum test with continuity correction

data:  employment_services by region
W = 5352, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0

>
```

Figure 29. Wilcoson Result

The Mann-Whitney-Wilcoxon Test produced a W statistic of 5352 and a p-value < 2.2e-16, indicating a substantial shift in median ranks for employment services between Europe and Sub-Saharan Africa. Strong evidence supports rejecting the null hypothesis, signifying a significant distribution difference in employment services.

4.0 Regression Analysis

To construct a reliable regression model, certain assumptions must be satisfied, encompassing linearity, normality, independence of residuals, absence of homoscedasticity, and no multicollinearity.

Regression Model 1 aims to predict the average employment service for the year 2016 based on relevant attributes.

```r
##    REGRESSION ANALYSIS FOR THE INDICATORS          ####

# Sub setting the data for the year 2016
tr <- jobs[jobs$year == "2016",]

# Creating a pairs plot for selected columns
pairs(tr[, c(3:10)], lower.panel = NULL, pch = 19, cex = 0.2)

# Removing unnecessary columns
jb <- tr[, -which(names(tr) %in% c('year', 'country', 'region'))]

# Checking the model to see if there's linearity
pairs(jb[, c(1:8)], lower.panel = NULL, pch = 19, cex = 0.2)

# Calculating Spearman correlation for the 'jb' dataset
cor_matrix <- cor(jb, method = 'spearman')

# Plotting the correlation matrix
corrplot(cor_matrix, type = "lower",
         tl.col = "black",
         method = 'number',
         tl.srt = 45)
```

Figure 30. Pair Scatter Plot

The subsequent step involves creating a forward stepwise model, with further details and outcomes explained in the following text.

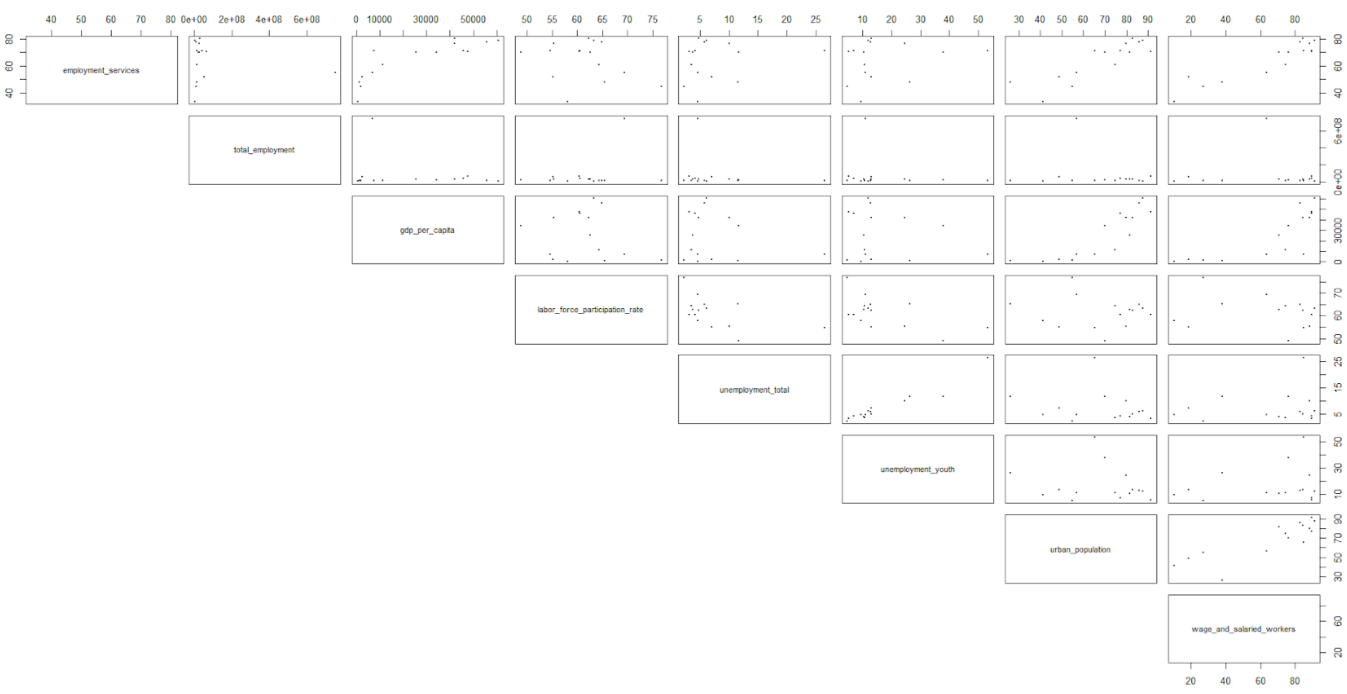


Figure 31. Pair Scatter Plot

The scatter plot in Figure 31 indicates a linear association between employment service with GDP per capita, urban population, and wage and salaried workers. The subsequent step involves creating a forward stepwise model, with further details and outcomes explained in the following text.

Figure 32. Correlation Analysis

Figure 32 indicates a strong correlation between employment service and the wage and salaried worker in the year 2016. Consequently, the wage and salaried worker will serve as the initial regressor in the equation.

```
# Displaying  summary of the regression model
summary(model1)

# Checking for Linearity
crPlots(model1)

# Checking for Normality of Residuals
plot(model1, which = 1)

# Checking for Independence of Residuals
plot(model1, which = 2)

# Checking for Homoscedasticity
plot(model1, which = 3)

```

Figure 33. Code

```
> model1 <- lm(employment_services ~ wage_and_salaried_workers, data = jb)
> model1

Call:
lm(formula = employment_services ~ wage_and_salaried_workers,
    data = jb)

Coefficients:
              (Intercept)  wage_and_salaried_workers
                   32.922                      0.473

>
> # Displaying  summary of the regression model
> summary(model1)

Call:
lm(formula = employment_services ~ wage_and_salaried_workers,
    data = jb)

Residuals:
    Min      1Q  Median      3Q     Max
-8.1647 -4.1393 -0.8701  3.4115 10.1871

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                32.9216     3.7836   8.701 8.81e-07 ***
wage_and_salaried_workers   0.4730     0.0531   8.909 6.76e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.596 on 13 degrees of freedom
Multiple R-squared:  0.8593,	Adjusted R-squared:  0.8484
F-statistic: 79.36 on 1 and 13 DF,  p-value: 6.758e-07
```

Figure 34. Output

The wage and salaried workers coefficient is highly significant ($p < 0.001$), implying a positive association. The R-squared (0.86) indicates that around 86% of employment services variability is explained by wage and salaried workers. The intercept at 32.92 estimates employment services when workers are zero. The F-statistic (79.36) and p-value (6.76e-07) confirm the overall model significance.

Checking for assumptions

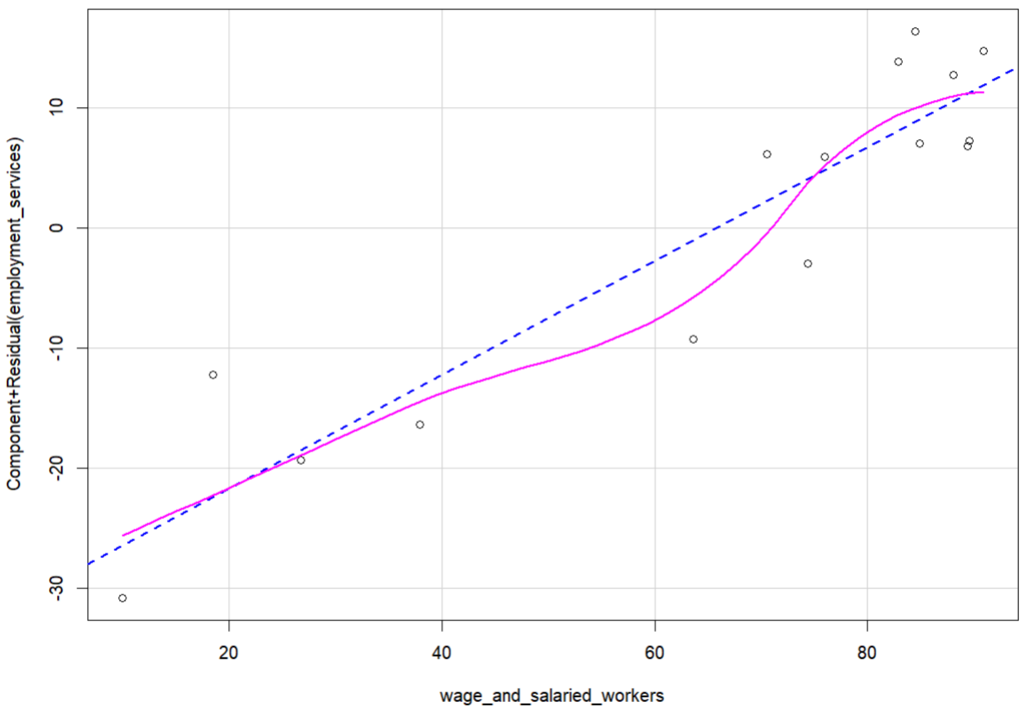


Figure 35. Regression line

In Figure 35, the association between the predictor variable is nearly linear, showing a relationship between the predictor variable.

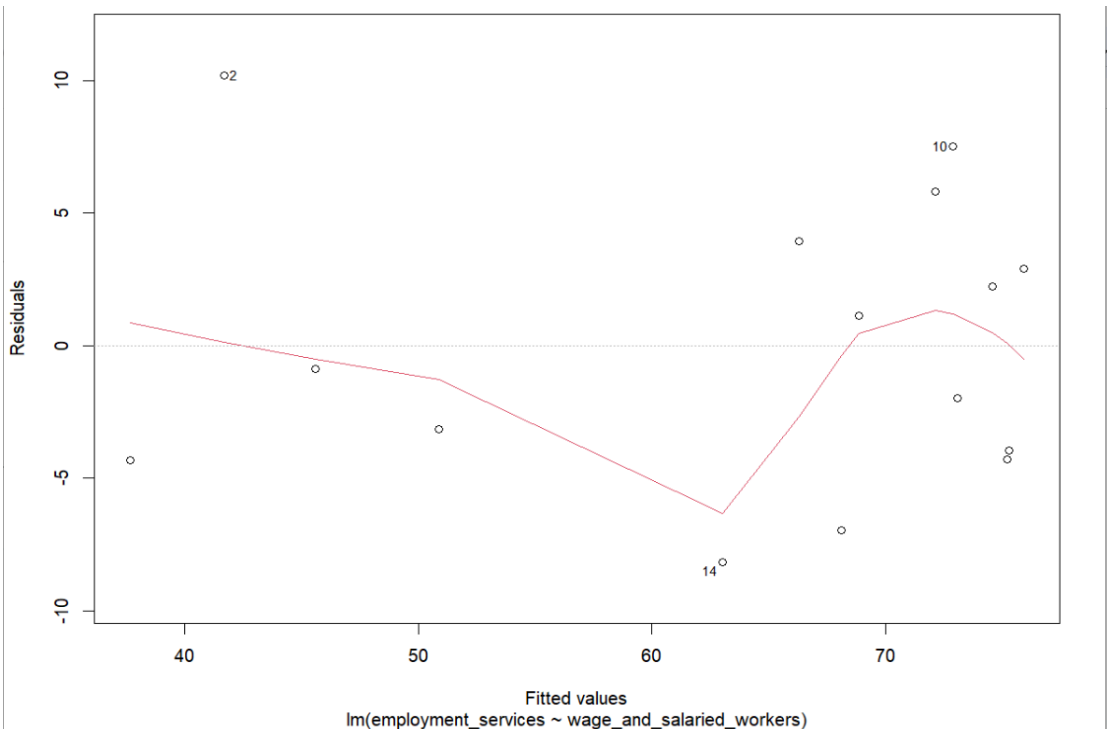


Figure 36. Normality of Residuals

In Figure 36, the residuals' distribution points appear to be close to a normal distribution.

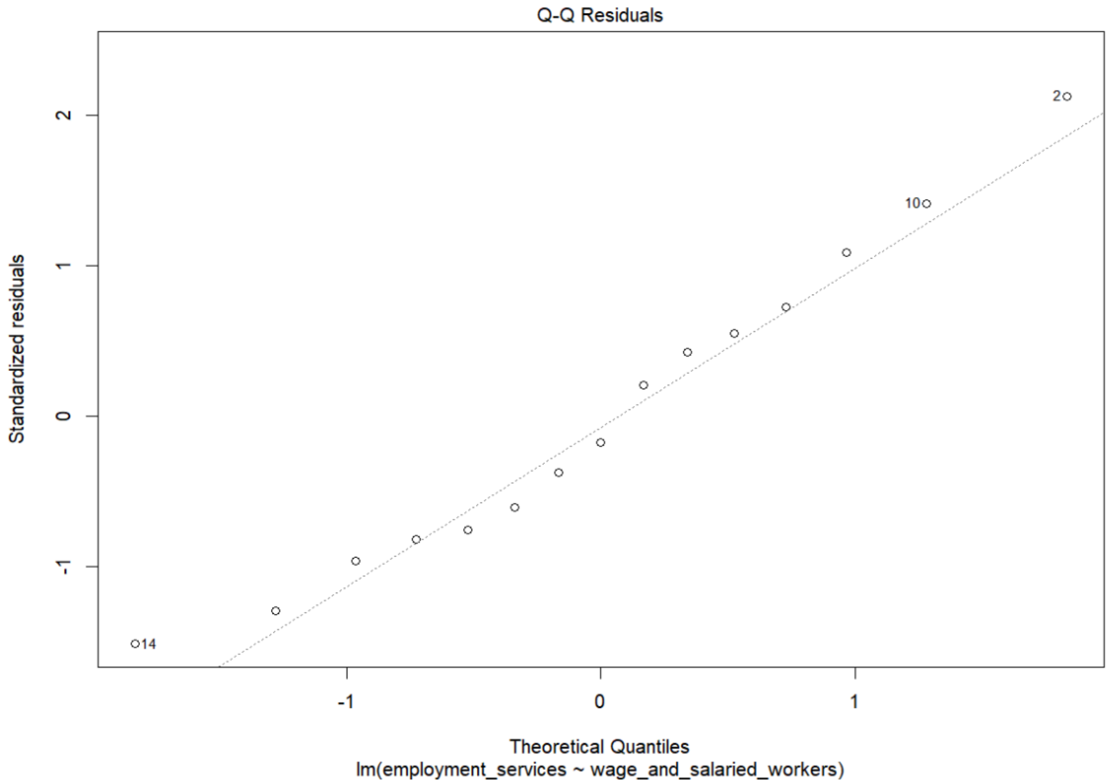


Figure 37. Independence of residuals

In Figure 37, the residuals are independent and display a nearly normal distribution.

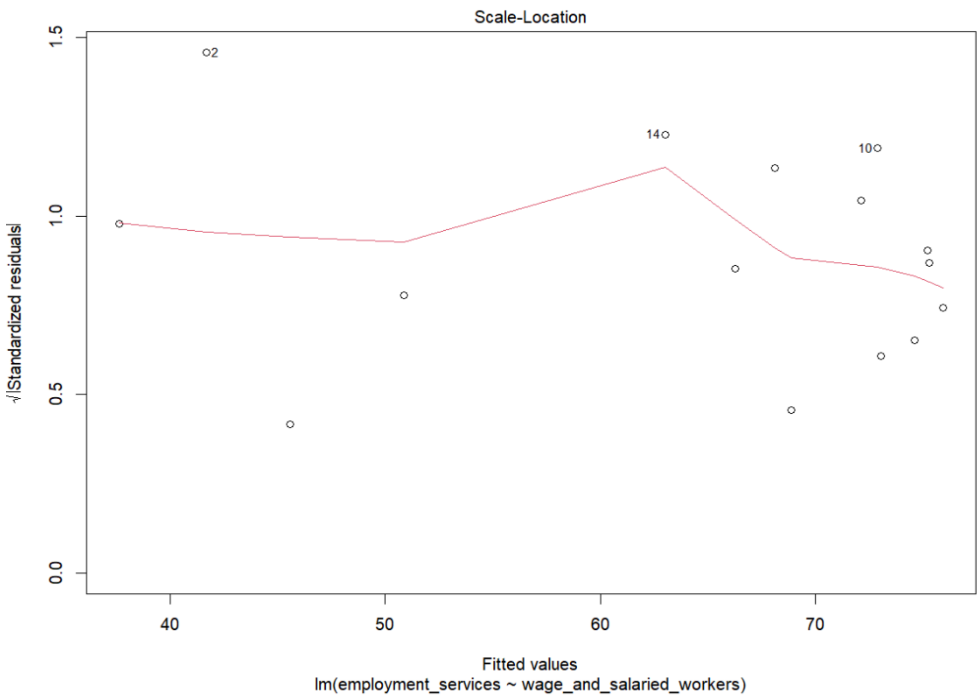


Figure 38. Homoscedasticity

Figure 38 exhibits consistent variability across various levels of the predictor, forming an almost horizontal line.

All assumptions have been satisfied. In conclusion, this model is deemed effective for predicting employment services considering the specified factors.

The fitted regression equation for Employment Services Model 1 is

$$\text{Employment services} = \beta 0 + \beta 1 \text{ X urban\_population}$$

$$\text{Employment services} = 32.9216 + 0.4730 \times \text{wage\_and\_salaried\_workers}$$

Assessing the model's fit by comparing its predictions with the actual values using the provided equation.

Wage and salaried workers	Actual	Predicted
84.849	71.088	73.05708
26.801	44.729	45.59905
37.918	47.710	50.85765

Table 10. Model Testing

4.1 Multiple Linear Regression Model and Comparison with Model 1:

Comparison of Model 2 Multiple Regression with Model 1:

In the stepwise multiple regression, GDP per capita was added, boosting R-squared from 0.84 to 0.896. Residual standard errors dropped from 5.596 to 5.007, with a p-value below 0.05. Model 2 shows a higher Multiple R-squared, indicating enhanced explanatory power. Both models are statistically significant and perform well in explaining the data. (See Figure 39 & 40)

```r
## MULTIPLE LINEAR REGRESSION
# Using lm() for performing multiple linear regression analysis
model <- lm(employment_services ~ wage_and_salaried_workers + gdp_per_capita, data = jb)
model

# Checking for multicollinearity using VIF
vif_values <- car::vif(model1)
print(vif_values)

# Displaying summary of the regression model
summary(model)

# Checking for Linearity
crPlots(model)

# Checking for Normality of Residuals
plot(model, which = 1)

# Checking for Independence of Residuals
plot(model, which = 2)

# Checking for Homoscedasticity
plot(model, which = 3)
```

Figure 39. Code

```
> ## MULTIPLE LINEAR REGRESSION
> # Using lm() for performing multiple linear regression analysis
> model <- lm(employment_services ~ wage_and_salaried_workers + gdp_per_capita, data = jb)
> model

Call:
lm(formula = employment_services ~ wage_and_salaried_workers + 
    gdp_per_capita, data = jb)

Coefficients:
              (Intercept)  wage_and_salaried_workers             gdp_per_capita  
                3.590e+01                  3.500e-01                  1.997e-04  

> 
> # Displaying summary of the regression model
> summary(model)

Call:
lm(formula = employment_services ~ wage_and_salaried_workers + 
    gdp_per_capita, data = jb)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.1780 -3.8364 -0.8908  2.9505  8.9965 

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)               3.590e+01  3.682e+00   9.750  4.7e-07 ***
wage_and_salaried_workers 3.500e-01  7.636e-02   4.583 0.000629 ***
gdp_per_capita            1.997e-04  9.704e-05   2.058 0.061986 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.007 on 12 degrees of freedom
Multiple R-squared:  0.896,	Adjusted R-squared:  0.8786 
F-statistic: 51.68 on 2 and 12 DF,  p-value: 1.268e-06
```

Figure 40. Output

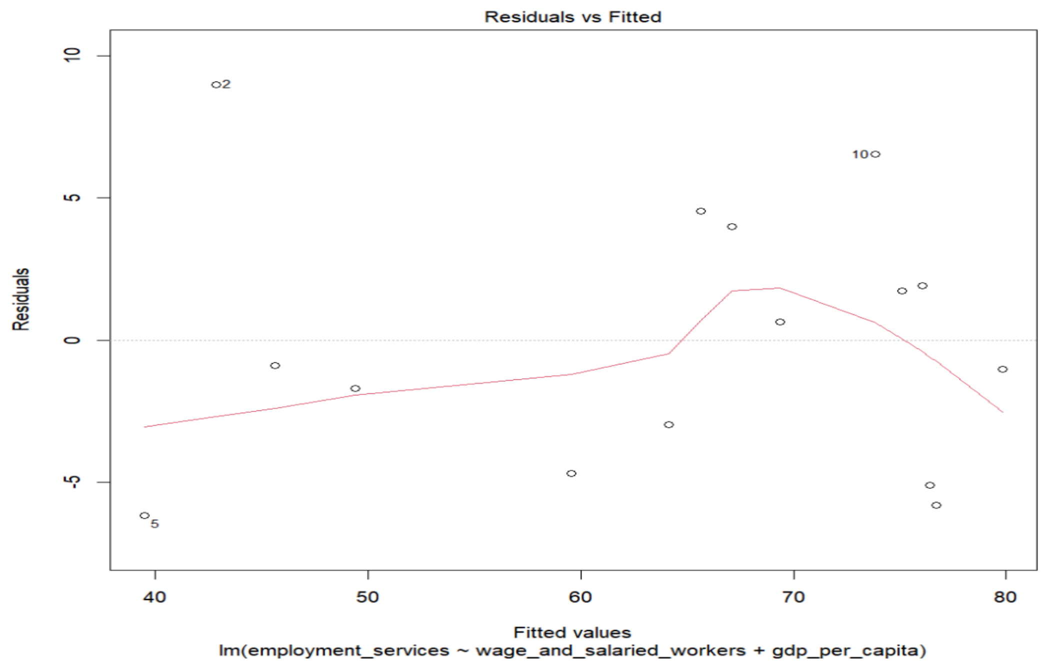


Figure 41. Residual Independence

From Figure 41, an almost flat line signifies no correlation among residuals

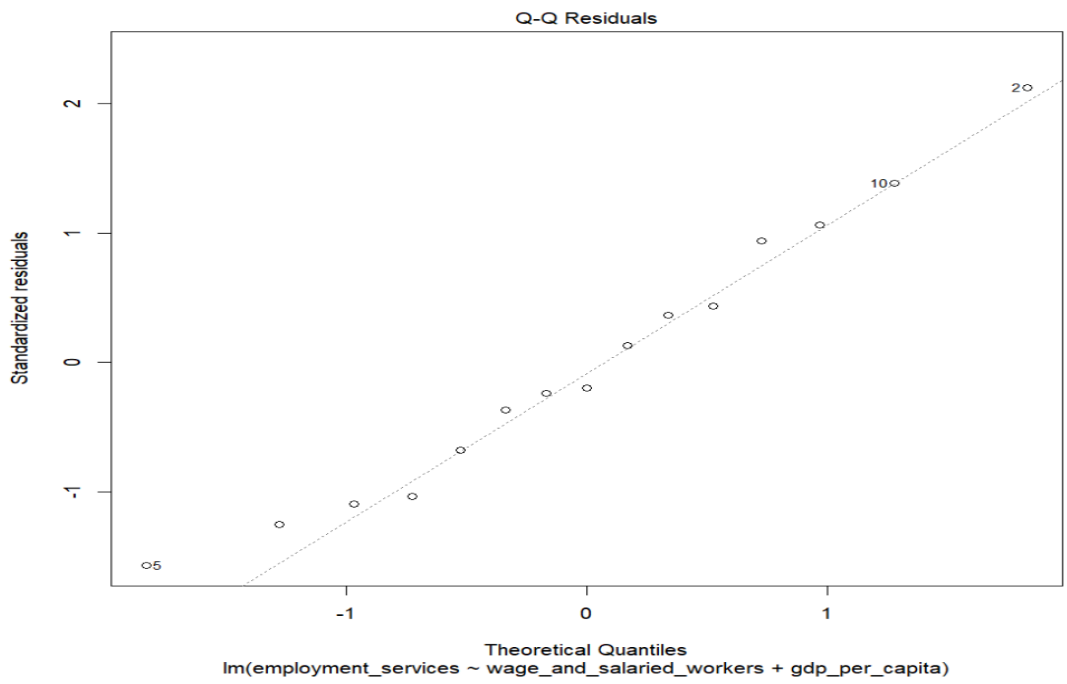


Figure 42. Normal Q-Q Residuals

From Figure 42, the majority of data points align with the line, suggesting normality.

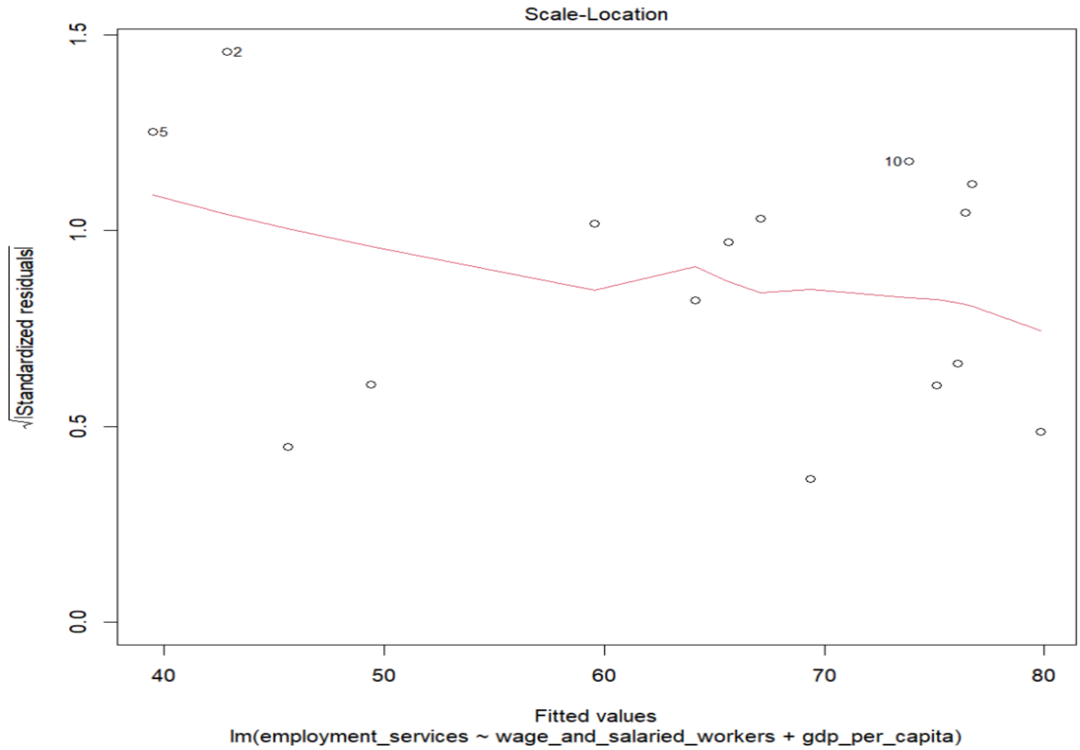


Figure 43. Fitted Residual Values

Figure 44, uniform variability is observed among fitted values, and considering the limited number of observations, the graph is deemed acceptable.

```
>
> # Checking for multicollinearity using VIF
> vif_values <- car::vif(model)
> print(vif_values)
wage_and_salaried_workers            gdp_per_capita
                 2.583165                  2.583165
>
```

Figure 45. Multicollinearity.

As observed in Figure 46, the variance inflation factor (VIF) values for both predictors in Model 2, with a value of 2.58 which is > 5, indicate the absence of substantial multicollinearity.

Considering all assumption has been met, the fitted regression equation is

$$Employment\_Services=35.90+0.3500\times(Wage\_and\_Salaried\_Workers)+0.0001997\times(GDP\_per\_Capita)$$

5.0 Time Series Analysis

For the fourth objective, construction and evaluation of two prevalent time series models which are Holt-Winters and ARIMA time series to forecast Employment services across three distinct regions. The primary objective is to compare the performance of these models in predicting future trends and the variations in Employment services within each region. providing valuable insights into the temporal patterns and behaviours of this critical economic indicator.

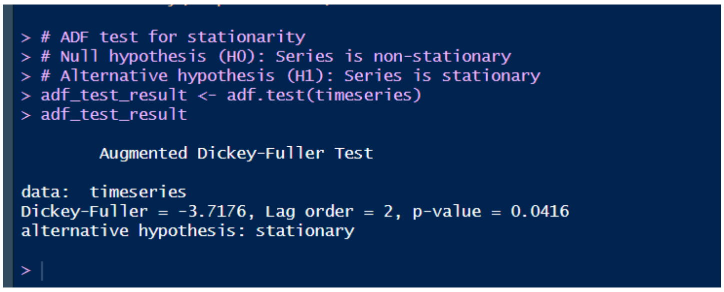
```r
##   TIME SERIES ANALYSIS FOR EMPLOYMENT SITUATION      ####
library(TTR)
library(forecast)
library(tseries)

# Defining a new ES_ts - Employment Situation time series
ES_ts <- each_country_mean[c('employment_services_mean')]
glimpse(ES_ts)

# Employment Situation time series
timeseries <- ts(ES_ts, start = c(2002))
timeseries
plot.ts(timeseries)

acf(timeseries)

# Print the characteristics of the time series
print(timeseries)

# ADF test for stationarity
# Null hypothesis (H0): Series is non-stationary
# Alternative hypothesis (H1): Series is stationary
adf_test_result <- adf.test(timeseries)
adf_test_result
```

Figure 47. Code Snippet for time series

```
> # Defining a new ES_ts - Employment Situation time series
> ES_ts <- each_country_mean[c('employment_services_mean')]
> glimpse(ES_ts)
Rows: 15
Columns: 1
$ employment_services_mean <dbl> 75.96987, 42.14673, 75.87213, 74.10880, 68.91800, 39.94453, 67.00753, 68.78753, 45.2…
>
> # Employment Situation time series
> timeseries <- ts(ES_ts, start = c(2002))
> timeseries
Time Series:
Start = 2002
End = 2016
Frequency = 1
      employment_services_mean
 [1,]                 75.96987
 [2,]                 42.14673
 [3,]                 75.87213
 [4,]                 74.10880
 [5,]                 68.91800
 [6,]                 39.94453
 [7,]                 67.00753
 [8,]                 68.78753
 [9,]                 45.23093
[10,]                 67.39747
[11,]                 57.40760
[12,]                 45.99793
[13,]                 30.57527
[14,]                 68.35360
[15,]                 78.10487
```

Figure 48. Time Series Ouput

The mean values for employment services suggest a fluctuating trend, indicating variations over time.



```
> # ADF test for stationarity
> # Null hypothesis (H0): Series is non-stationary
> # Alternative hypothesis (H1): Series is stationary
> adf_test_result <- adf.test(timeseries)
> adf_test_result

        Augmented Dickey-Fuller Test

data:  timeseries
Dickey-Fuller = -3.7176, Lag order = 2, p-value = 0.0416
alternative hypothesis: stationary

>
```

Figure 49. ADF Test

The Augmented Dickey-Fuller (ADF) test on the time-series data yields a test statistic of -3.7176, a lag order of 2, and a p-value of 0.0416. The result suggests rejecting the null hypothesis of non-stationarity, indicating that the series is likely stationary.

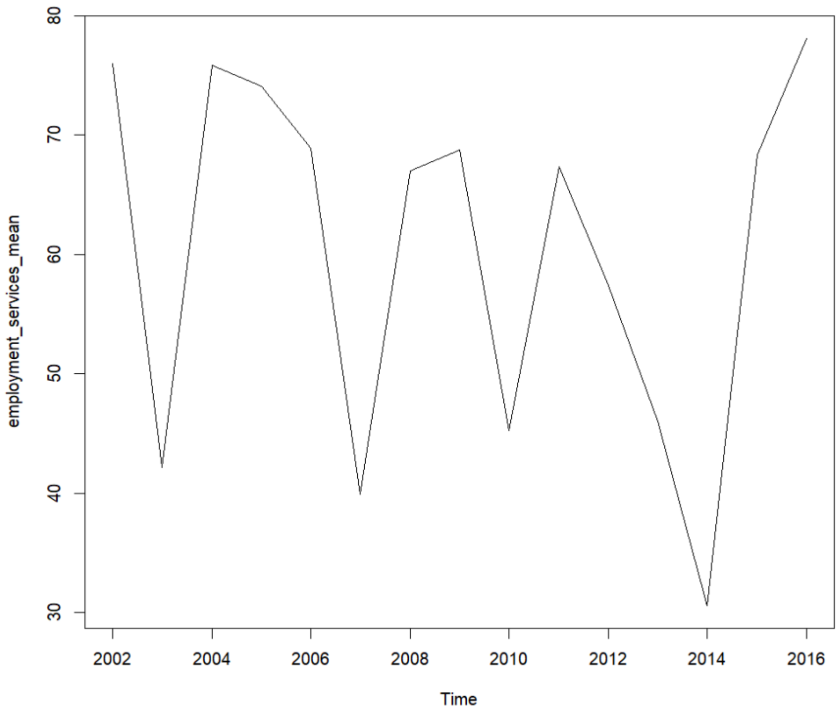


Figure 50. Time series plot

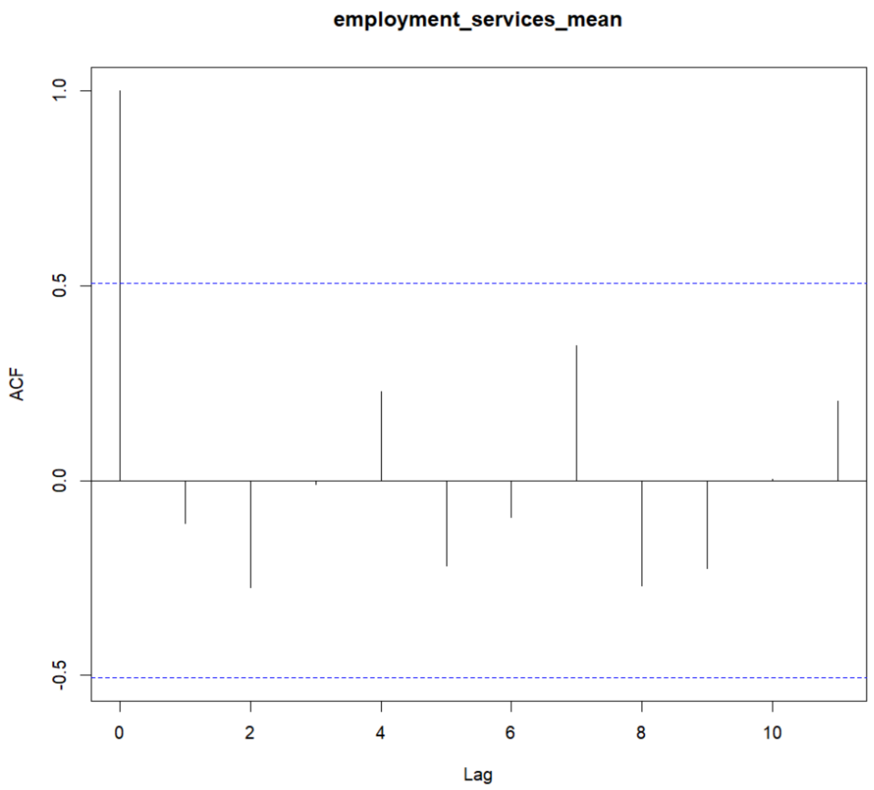


Figure 51. ACF plot

5.1 Holt-Winters Time Series Model

In time series analysis, the Holt-Winters method utilizes three parameters: alpha, beta, and gamma, which determine the coefficients for smoothing the level, trend, and seasonality, respectively. (Holt-Winters(Time Series), n.d.). Gamma is excluded as this series is considered stationary.

```r
## 	MODEL1 HOLT-WINTERS SERIES 			#####
ES_forecast <- HoltWinters(timeseries, gamma = FALSE)
ES_forecast
ES_forecast$fitted
plot(ES_forecast)

# SSE - sum of squared errors
ES_forecast$SSE

# Shift the graph up
plot(HoltWinters(timeseries, gamma = FALSE))

require("forecast")
# Forecasting for 8 steps ahead
ES_forecast2 <- forecast(ES_forecast, h = 8)
ES_forecast2
plot(ES_forecast2)
ES_forecast$fitted
ES_forecast2$x

# Autocorrelation plot of residuals
acf(ES_forecast2$residuals, lag.max = 4, na.action = na.pass)

# Partial Autocorrelation plot of residuals
pacf(ES_forecast2$residuals, lag.max = 4, na.action = na.pass)

# Plotting Residuals against time
plot(ES_forecast2$residuals, ylab = "ES_forecast2$residuals", xlab = "Time")

# Plotting Residuals: Forecast errors plotted on the time graph
plot.ts(ES_forecast2$residuals)

# Ljung-Box test for autocorrelation of residuals
Box.test(ES_forecast2$residuals, lag = 4, type = "Ljung-Box")


plotForecastErrors <- function(forecasterrors)
{…}

plotForecastErrors(ES_forecast2$residuals)
```

```
> ES_forecast
Holt-Winters exponential smoothing with trend and without seasonal component.

Call:
HoltWinters(x = timeseries, gamma = FALSE)

Smoothing parameters:
 alpha: 0.8279835
 beta : 0.3612592
 gamma: FALSE

Coefficients:
       [,1]
a 75.966277
b  8.563906
>
```

Figure 52. Code Snippet and result for HoltWinters

To preserve degrees of freedom, Jeffery Wooldridge suggests choosing a small number of lags, typically 1 or 2. For non-seasonal data, a total lag of 4 is recommended (Wooldridge, 2002).

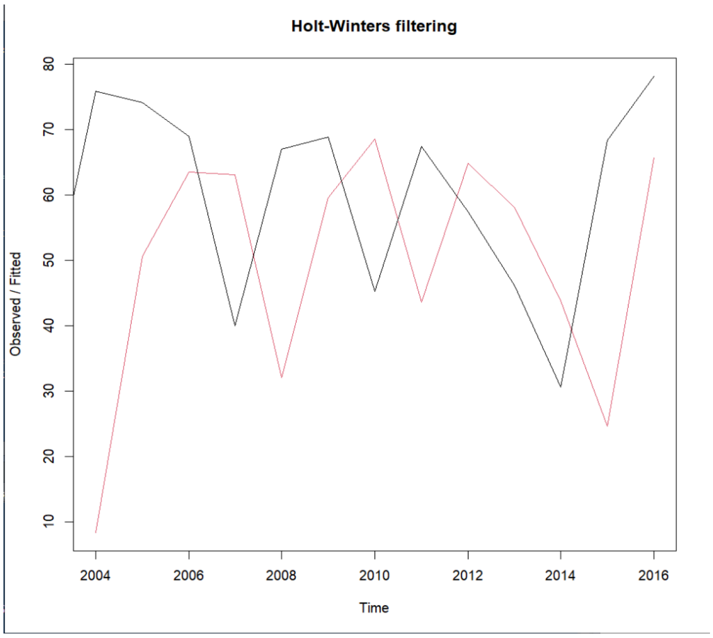


Figure 53. Holt-Winters Fittering

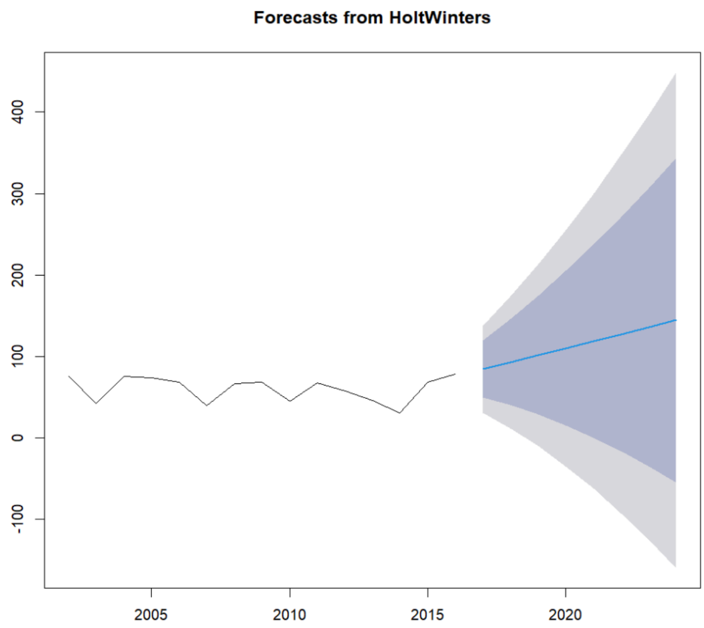


Figure 54. Holt-Winters Forecast

```
> require("forecast")
> # Forecasting for 8 steps ahead
> ES_forecast2 <- forecast(ES_forecast, h = 8)
> ES_forecast2
     Point Forecast        Lo 80    Hi 80       Lo 95    Hi 95
2017       84.53018  49.4575531 119.6028   30.891225 138.1691
2018       93.09409  40.2477153 145.9405   12.272536 173.9156
2019      101.65799  28.8922269 174.4238   -9.627643 212.9436
2020      110.22190  15.5828138 204.8610  -34.516092 254.9599
2021      118.78580   0.4728628 237.0987  -62.158225 299.7298
2022      127.34971 -16.3136396 271.0131  -92.364423 347.0638
2023      135.91362 -34.6744499 306.5017 -124.978317 396.8055
2024      144.47752 -54.5236256 343.4787 -159.868469 448.8235
```

Figure 55. Forecast Prediction

Forecast predicts an increasing trend in employment (84.53 to 144.48).

```
> # Ljung-Box test for autocorrelation of residuals
> Box.test(ES_forecast2$residuals, lag = 4, type = "Ljung-Box")

        Box-Ljung test

data:  ES_forecast2$residuals
X-squared = 1.6441, df = 4, p-value = 0.8008
```

Figure 56. Box-Ljung test

The residuals of the forecast indicate a lack of evidence for autocorrelation, with a p-value of 0.8008.

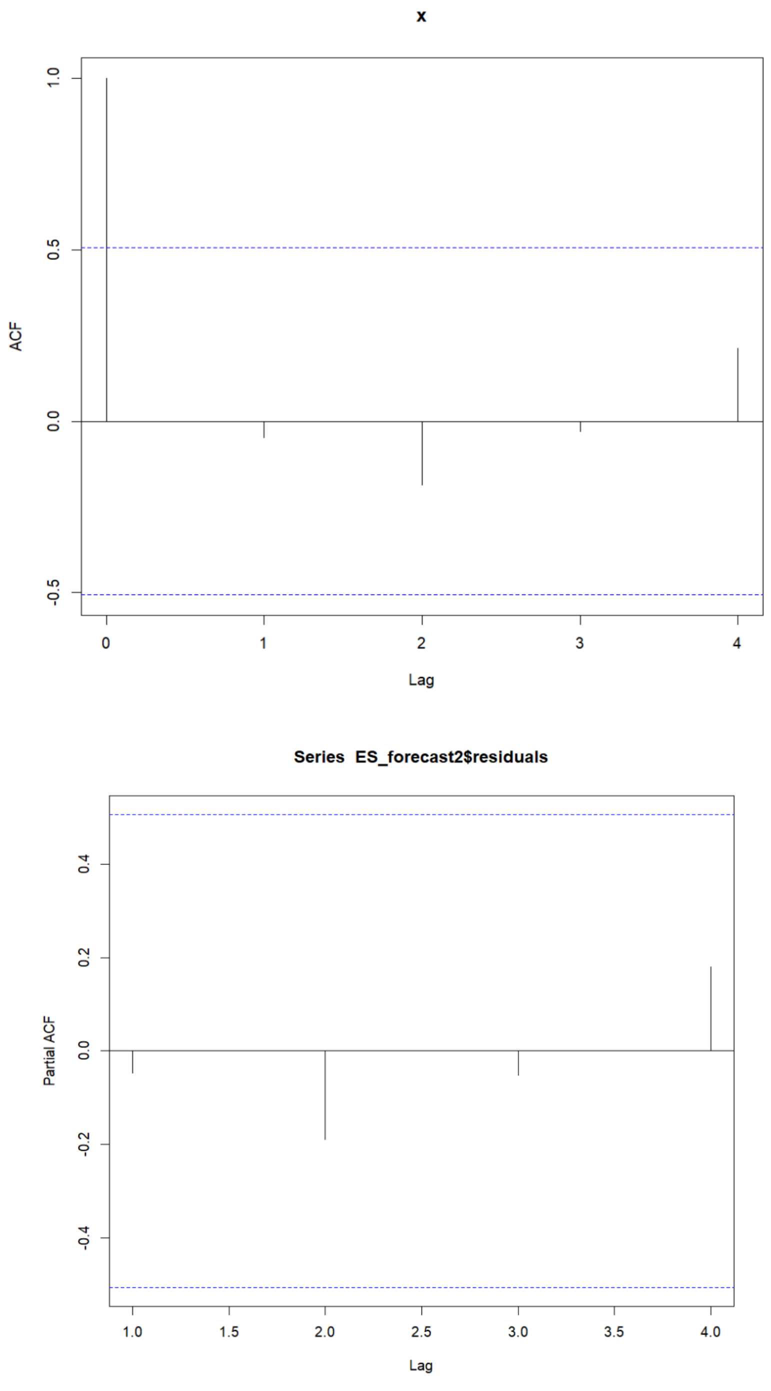


Figure 57. ACF and PACF

The autocorrelations of forecast errors in Figure 57 stay within significance bounds, suggesting a well-performing model without autocorrelation.

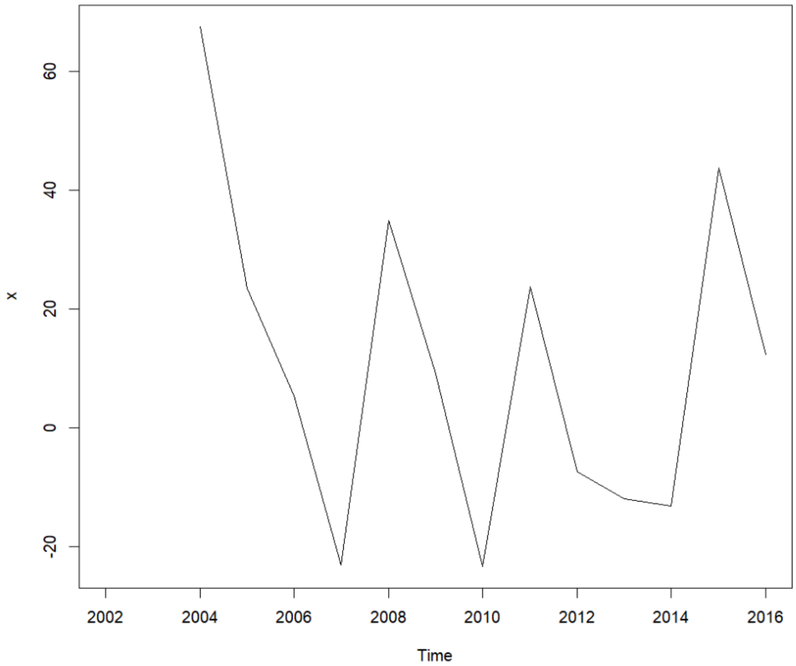


Figure 58. Residual Plot

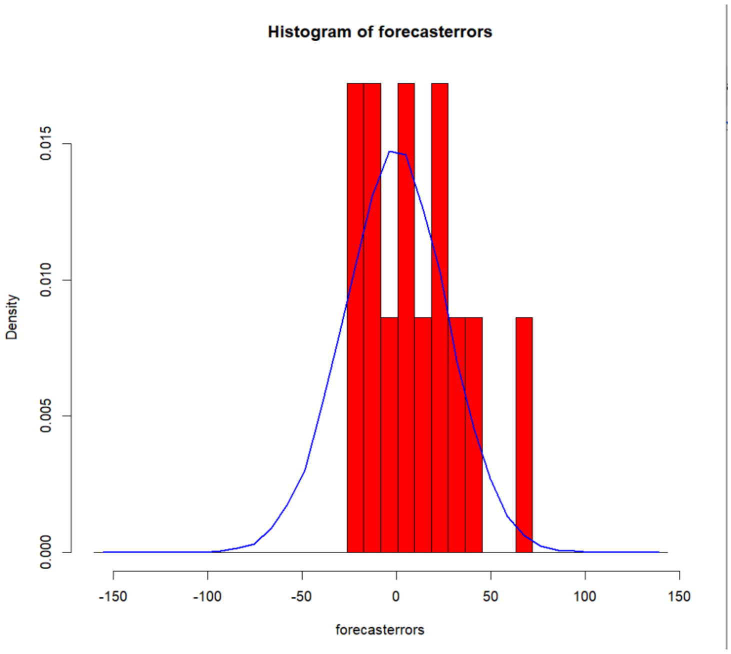


Figure 59. The Histogram of Forecasterrors

Residuals plotted on the time graph exhibits consistent variance with a subtle left skew, yet mostly conforms to normal distribution around mean 0 constant over time.

5.2 ARIMA Time Series Model

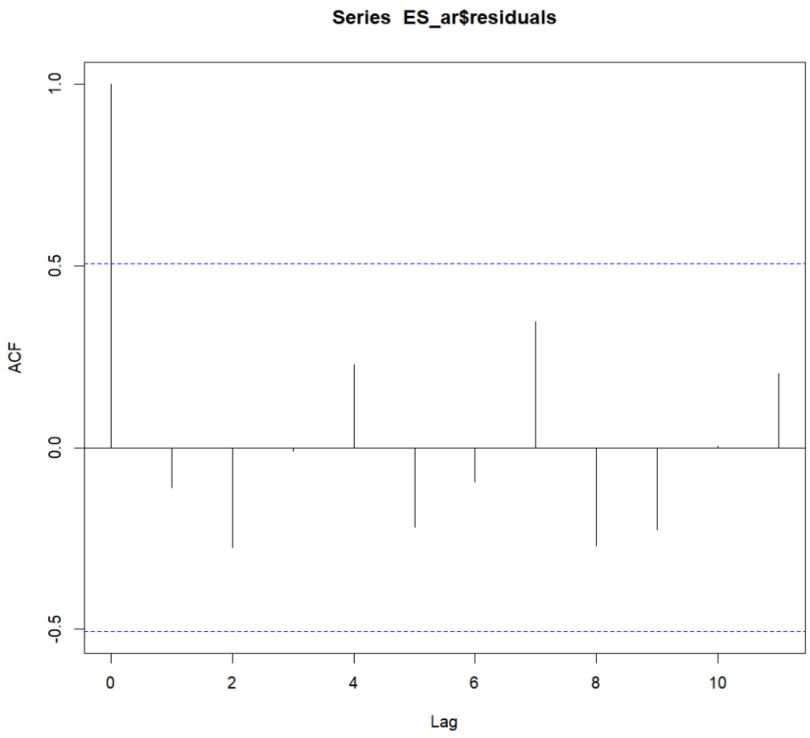
Arima assumes stationarity, and the ADF test in Figure 49 indicates Stationarity. There will be no need to differentiate the series, but a check can still be done.

```
##        MODEL2 ARIMA TIME SERIES                              ####
# Load necessary libraries

# Display the time series
timeseries

# Check ACF plot for autocorrelation
acf(timeseries)

# Fit ARIMA model
ES_ar <- auto.arima(timeseries)
ES_ar

# Autocorrelation plot of residuals
acf(ES_ar$residuals)

# Partial Autocorrelation plot of residuals
pacf(ES_ar$residuals)

# Plotting residuals
plot.ts(ES_ar$residuals)

# Forecasting
ES_ar_forecast <- forecast(ES_ar, h = 5)
ES_ar_forecast

# Ljung-Box test for autocorrelation of residuals
Box.test(ES_ar$residuals, type = "Ljung-Box")

# Plotting forecast
plot(ES_ar_forecast)

# Plotting forecast errors
plotForecastErrors(ES_ar_forecast$residuals)

# Filtering out NA values from forecast errors
filtered_forecast_errors <- ES_ar_forecast$residuals[!is.na(ES_ar_forecast$residuals)]

## Plotting filtered forecast errors
plotForecastErrors(filtered_forecast_errors)
```

```
>
> # Check ACF plot for autocorrelation
> acf(timeseries)
>
> # Fit ARIMA model
> ES_ar <- auto.arima(timeseries)
> ES_ar
Series: timeseries
ARIMA(0,0,0) with non-zero mean

Coefficients:
         mean
      60.3882
s.e.   3.8699

sigma^2 = 240.7:  log likelihood = -61.89
AIC=127.79   AICc=128.79   BIC=129.2
>
```

Figure 59. Code Snippet and result for ARIMA

Auto Arima has calculated values for p, d, and q, as indicated with non-zero mean



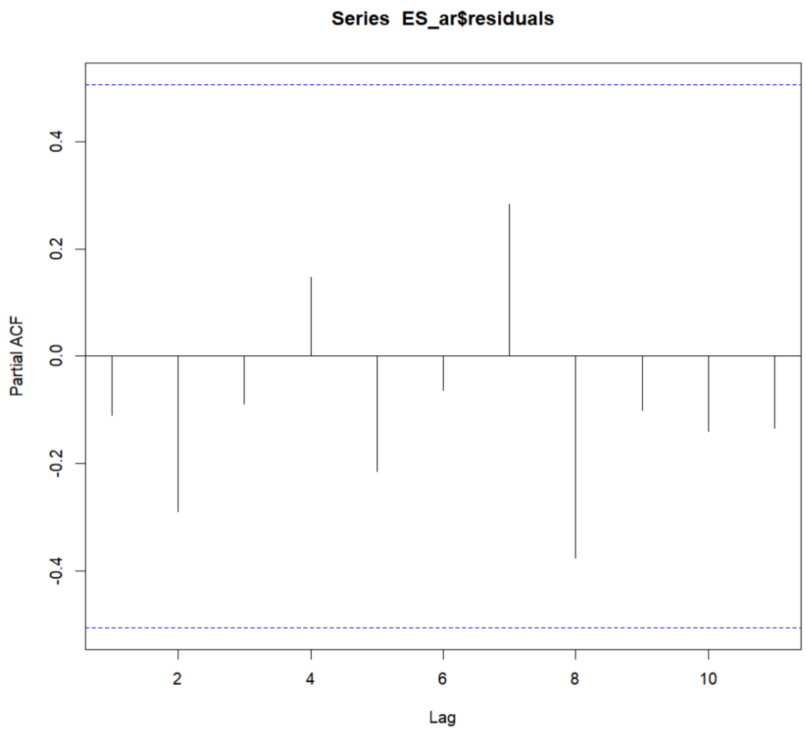


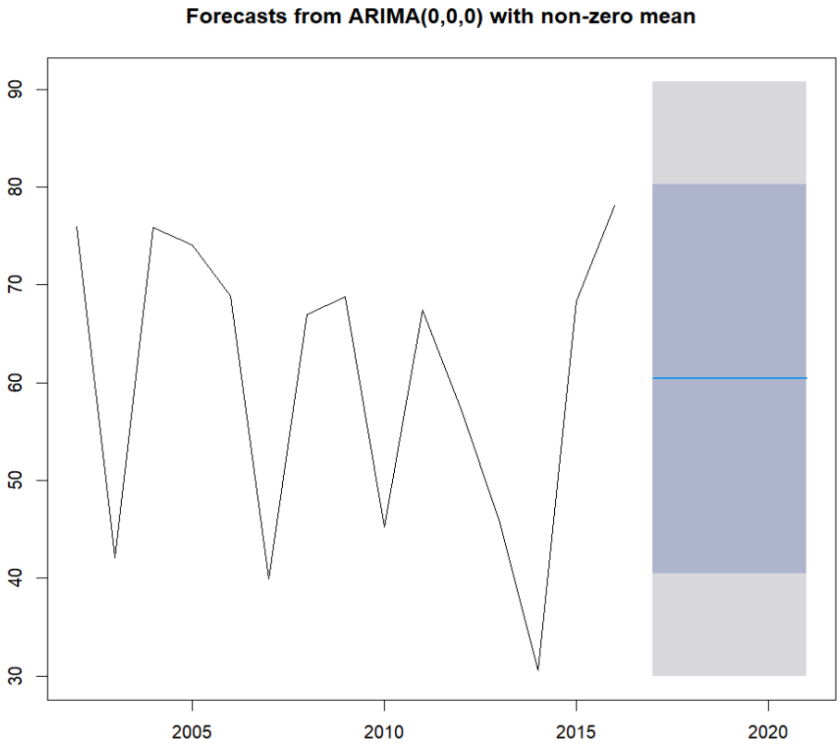
Figure 60. ACF and PACF

ACF and PACF, depicted in a correlogram, reveal no significant correlations among lagged values or forecasts.

```
> # Ljung-Box test for autocorrelation of residuals
> Box.test(ES_ar$residuals, type = "Ljung-Box")

        Box-Ljung test

data:  ES_ar$residuals
X-squared = 0.21984, df = 1, p-value = 0.6392
```

Figure 61. Box-Ljung Test

The Box test shows no significant autocorrelation in residuals ($p > 0.05$, X-squared = 0.2198), indicating uncorrelated errors over time.




```
> ES_ar_forecast
     Point Forecast    Lo 80   Hi 80    Lo 95    Hi 95
2017       60.38819 40.50617 80.2702 29.98127 90.79511
2018       60.38819 40.50617 80.2702 29.98127 90.79511
2019       60.38819 40.50617 80.2702 29.98127 90.79511
2020       60.38819 40.50617 80.2702 29.98127 90.79511
2021       60.38819 40.50617 80.2702 29.98127 90.79511
```

Figure 62. ARIMA Forecast

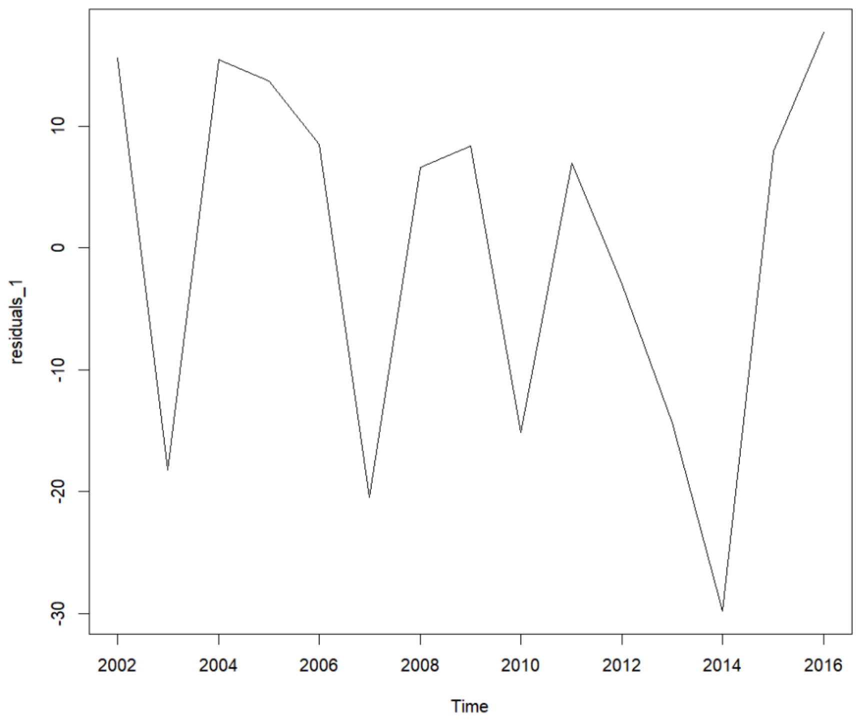


Figure 63. Residuals Time Plot

Residual plots in figure 63 above suggest almost stationary behaviour.

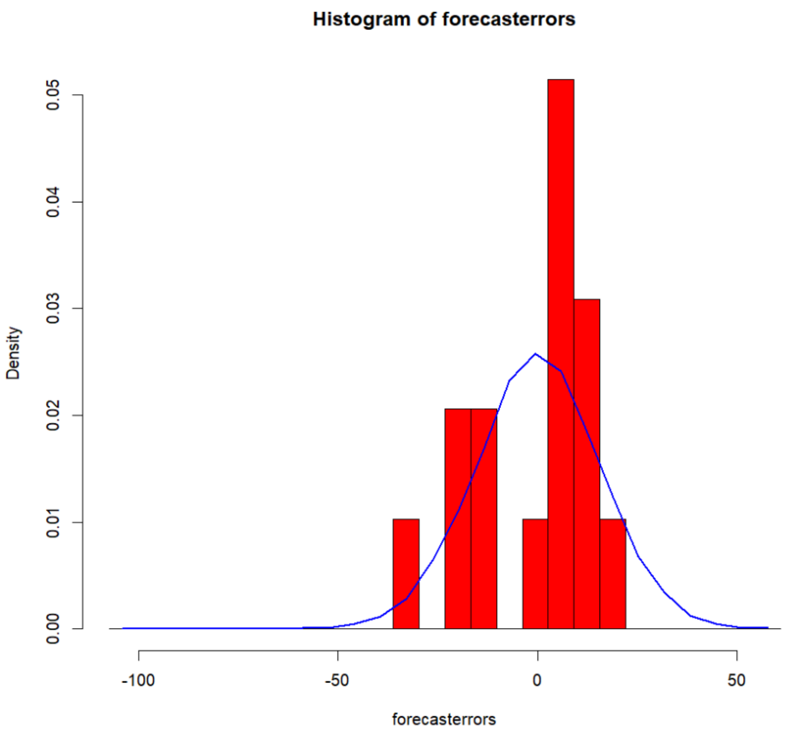


Figure 64. Forecast Errors

Errors follow a normal distribution around mean 0.

5.3 Comparing of Holt-Winter and ARIMA Models

The ARIMA model exhibits lower forecasting errors (MAE, MSE, RMSE) and better captures data variability compared to Holt-Winters. Higher MPE and MAPE in Holt-Winters indicate greater percentage deviation. In conclusion, ARIMA outperforms Holt-Winters for employment forecasting in services.

```
> # Comparing mean and accuracy for Holt-Winters
> accuracy_hw <- calculate_accuracy(actual_values, forecast_hw)
> cat("Holt-Winters Model Accuracy:\n")
Holt-Winters Model Accuracy:
> cat("ME:", accuracy_hw["ME"], "\n")
ME: -84.08933
> cat("MAE:", accuracy_hw["MAE"], "\n")
MAE: 84.08933
> cat("MSE:", accuracy_hw["MSE"], "\n")
MSE: 8856.673
> cat("RMSE:", accuracy_hw["RMSE"], "\n")
RMSE: 94.1099
> cat("MPE:", accuracy_hw["MPE"], "\n")
MPE: -162.8588
> cat("MAPE:", accuracy_hw["MAPE"], "\n")
MAPE: 162.8588
> cat("MASE:", accuracy_hw["MASE"], "\n")
MASE: 4.486575
> cat("ACF1:", accuracy_hw["ACF1"], "\n\n")
ACF1: 0.7125026

>
> # Comparing mean and accuracy for ARIMA
> accuracy_arima <- calculate_accuracy(actual_values, forecast_arima)
> cat("ARIMA Model Accuracy:\n")
ARIMA Model Accuracy:
> cat("ME:", accuracy_arima["ME"], "\n")
ME: 2.510573e-14
> cat("MAE:", accuracy_arima["MAE"], "\n")
MAE: 13.47015
> cat("MSE:", accuracy_arima["MSE"], "\n")
MSE: 224.6391
> cat("RMSE:", accuracy_arima["RMSE"], "\n")
RMSE: 14.98797
> cat("MPE:", accuracy_arima["MPE"], "\n")
MPE: -8.221353
> cat("MAPE:", accuracy_arima["MAPE"], "\n")
MAPE: 26.70597
> cat("MASE:", accuracy_arima["MASE"], "\n")
MASE: 0.718698
> cat("ACF1:", accuracy_arima["ACF1"], "\n")
```

Figure 65. Comparing two Models.

Discussion and Conclusion

6.0 Discussion

The histograms reveal the distribution shapes, with notable outliers in negatively skewed Employment Service distribution, suggesting countries with high employment. Similar trends are observed in other distributions like GDP per capita, urban population, wage and salaried workers, and labor force participation rate.

Mean and median analyses highlight regional disparities, with East Asia consistently having the highest values and Sub-Saharan Africa the lowest. This emphasizes the need for attention and interventions to address potential imbalances. Spearman's rank correlation matrix indicates strong positive correlations of employment services with urban population, GDP per capita, and wage and salaried workers, while labour force participation shows a weak negative correlation. This information is crucial for understanding the relationships between variables and potential implications for regional disparities. Density plots and hypothesis tests are utilized to assess the statistical significance of regional differences in employment services. The Kruskal-Wallis test indicates significant disparities among regions, and pairwise Wilcoxon tests confirm specific group differences. The Mann-Whitney-Wilcoxon Test further supports significant distribution differences between Asia and Europe.

Regression Model 1 predicts average employment services based on relevant attributes. The forward stepwise model, with wage and salaried workers as the initial regressor, demonstrates strong significance, explaining around 86% of the variability in employment services. Assumption checks confirm the model's reliability for predicting employment services. Multiple Linear Regression Model 2 includes GDP per capita, enhancing explanatory power with a higher R-squared. Assumption checks, including VIF for multicollinearity, validate the model's effectiveness in explaining data variability. Both Holt-Winters and ARIMA time series models are employed to forecast employment services. A stationary series is confirmed by the Augmented Dickey-Fuller test. While both models perform well, ARIMA outperforms Holt-Winters in terms of forecasting errors and capturing data variability. This information aids in understanding temporal patterns and behaviours of employment services within regions. The findings contribute to a better understanding of the factors influencing employment services across different regions.

Overall, the comprehensive statistical analysis provides valuable insights into the regional variations, relationships between variables, and predictive modelling for employment services.

7.0 Conclusion

In conclusion, the comprehensive statistical analysis reveals substantial regional variations in employment services, emphasizing disparities across East Asia, Europe, and Sub-Saharan Africa. Spearman's rank correlation exposes crucial relationships with urban population, GDP per capita, and wage and salaried workers. Regression models validated through thorough checks, provide effective predictors, with Model 2 incorporating GDP per capita for enhanced explanatory power. Time series analysis using Holt-Winters and ARIMA models forecasts employment services, with ARIMA demonstrating superior performance. Hypothesis testing confirms significant distribution differences among regions, highlighting the need for targeted interventions to address disparities. Overall, this analysis contributes valuable insights into regional employment dynamics and potential influencing factors.

Part Three: Interactive Dashboard Design

1.0 Investigation of Data Workflows & Proposal for Design of Dashboard

Data Attribute Characteristics:

In the chosen dataset, the categorical attributes encompass years and countries, while the remaining attributes take on continuous numerical values. The target variable is employment in the service sector, and it exhibits associations with variables such as wage and salaried workers, GDP per capita, urbanisation, population, etc. (Ma N et al, 2021). It's important to note that the study focuses on exploring relationships rather than establishing causation, and this distinction should be kept in mind.

Dashboard Design Steps Overview:

1. Selection of Attributes:
   - Chosen attributes based on research goals and objectives.
2. Data Retrieval:
   - Retrieved data from Databank.org, organized it with series as columns, and countries and time (year) as rows.
3. Data Cleaning and Analysis:
   - Utilized Power Query Editor to identify and eliminate missing values from the dataset.
4. Layout Planning:
   - Iteratively adjusted the layout using a trial-and-error approach until achieving the optimal arrangement aligned with research objectives.
5. Visualization Creation:
   - Employ Power BI to create insightful visuals for the dashboard.

Data preprocessing in Power Query:

After importing, data undergoes cleaning for effective utilization. This involves various procedures such as eliminating rows with substantial null values, adjusting data types, and arranging columns based on priority, with a key focus on essential indicators like Employment in services by country series. The process is illustrated in Figure 66

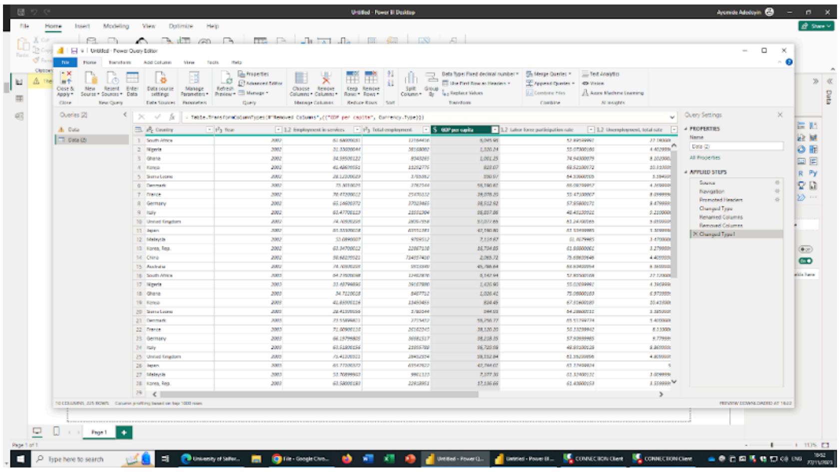


Figure 66. Power Query

The following are procedures taken in the power query editor

a. Adjusted Data Types:

The year was converted from integer to date format. The 'Country' categorical variable was retained as text, while other continuous variables were converted to decimal numbers.

b. Connection of Date/Year, Country and Region:

The slicer positioned at the dashboard's top establishes a link between the date or year and all other visuals on the dashboard.

c. Decimal Places:

The numbers were rounded to enhance clarity: Employment services were rounded to one decimal place, while other attributes were rounded to two decimal places for simplicity. This choice was justified due to significant differences in the numbers being compared.

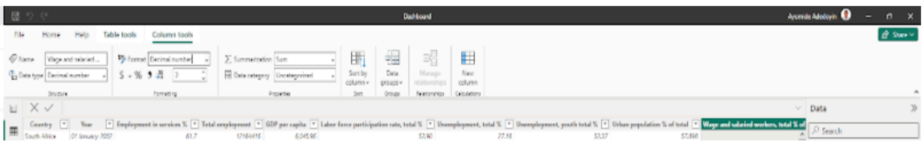


Figure 67. Decimal Places

d. Categorization Process:

In Figure 68, categorization was employed using the Power BI's 'grouping' functionality, countries were manually grouped into Europe and Africa based on location. Refer to Figures for details on the grouping process.

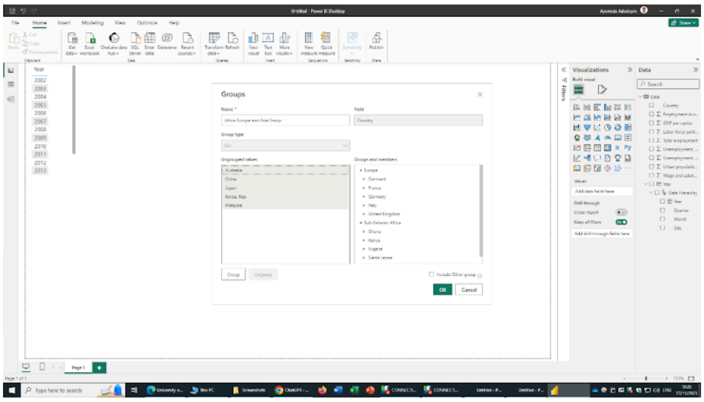


Figure 68. Country Grouping

e. Guidance for Visual Analysis:

Visual 1's tooltips in Figure 69 display crucial metrics influencing employment. Tooltips streamline information analysis within a glance, minimizing visual clutter (Wexler et al).

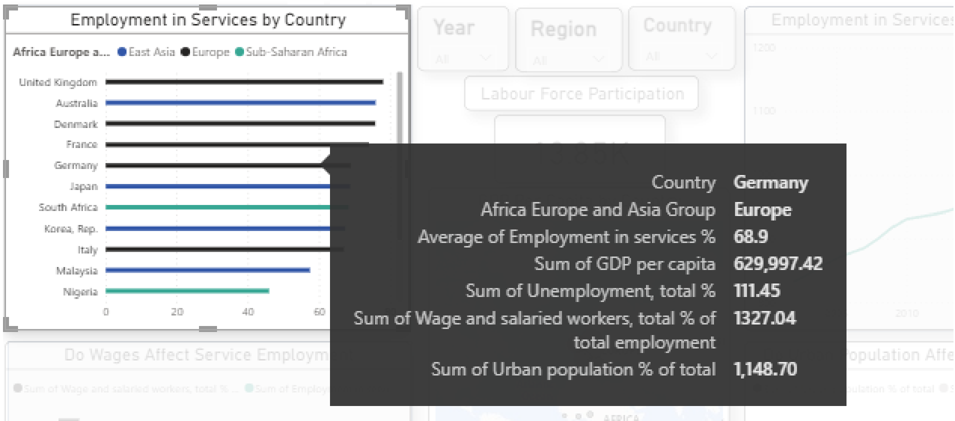


Figure 69. Tool Tips

Main Dashboard

There are a total of seven visuals. Positioned just below the title and above labour participation is the annual slicer visual, which serves to filter all other visuals on the dashboard for in-depth exploration.

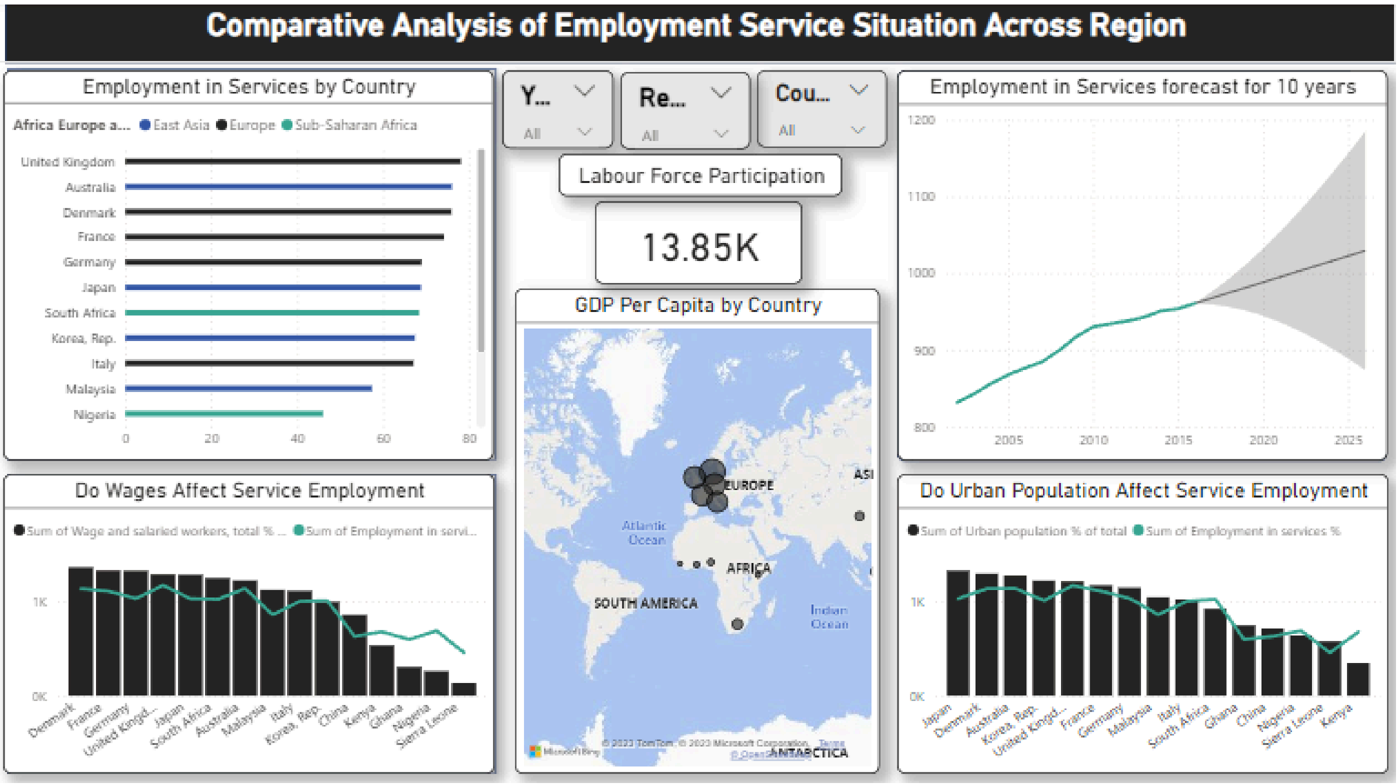


Figure 70. Dashboard

Overview of Dashboard

The top-left corner features summarising employment in services by country for Visual 1, followed by three slicers allowing data selection for year, country and region, which serves as Visual 2. The colour scheme, including white, cyan, grey, and black, adheres to Stephen Few's principle of colour as the initial perceptible element. Minimalist design is maintained with minimal borders, a shadow effect for a three-dimensional look, and rounded corners for aesthetic appeal and smooth transitions between visuals (Tseng, 2011)

Visuals 2 and 6 constitute combo charts, which amalgamate bar or cluster charts with a line chart. This type of chart is effective in illustrating the correlation between two variables (Visualization types in Power BI, 2022). It's important to observe the dual scales on each side, each corresponding to the attributes in the legend. The subsequent section will provide a detailed explanation of each visual, outlining the rationale for its selection and the content it conveys.

To minimize unnecessary noise and clutter devoid of new information, extraneous elements were eliminated. The title in the illustration has been condensed from "Average of employment services and the impact across the regions..." to "Does Wages Affect Service Employment?" with the aim of captivating users. Intriguing titles that evoke emotions have been shown to retain user engagement for extended periods (Arévalo & Devan, 2017).

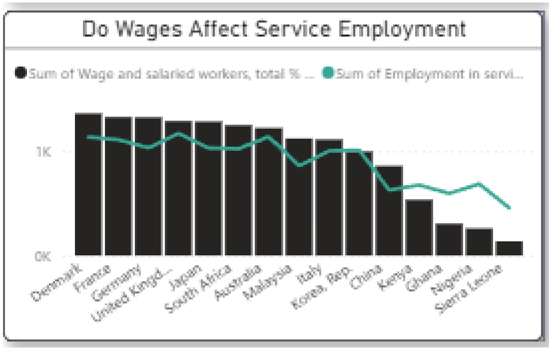


Figure 71. Simple Title

The pre-attentive attribute of visual encoding is employed to choose the black hue for emphasising the spectrum of developmental indicators across countries. According to Stephen Few's recommendations in his book "Dashboard Design for at-a-glance Monitoring" (2010), colour is the initial element that the human eye subconsciously detects.

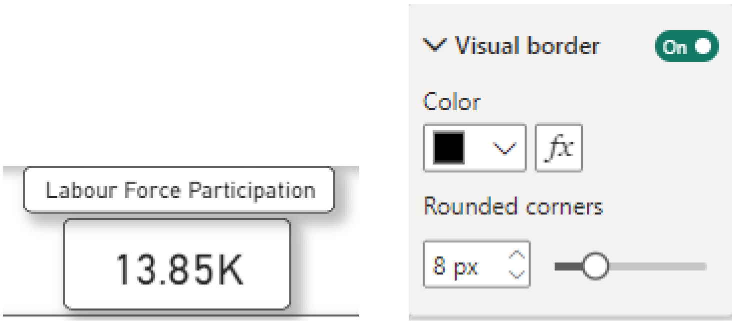


Figure 72. Visual Encoding

All graphics feature rounded corners to ensure a smooth flow. Antony, as stated by UX Movement, suggests that rounded corners offer visual appeal by reducing cognitive effort and ensuring a smooth transition for users between visuals without disrupting their reading flow (Tseng, 2011).

Visual 1: Employment in Services by Country

The visual in Figure 73 on Employment in Services by Country reveals insights into global economic trends. Countries like the United Kingdom lead in service sector employment, signalling robust economies. Meanwhile, regions in Sub-Saharan Africa face challenges with lower service sector development such as Sierra Leone. High numbers of wage workers in the UK and Australia showcase stable job markets, contrasting with some South Asian countries experiencing job instability. Urbanization is prominent in China and France, aligning with growing service activities, while rural areas in certain developing countries struggle with limited job creation. Sub=Saharan African countries face economic challenges. Low unemployment rates in Germany and Denmark reflect healthy service-based job markets, and some Sub-Saharan African nations grapple with high unemployment and limited service employment.

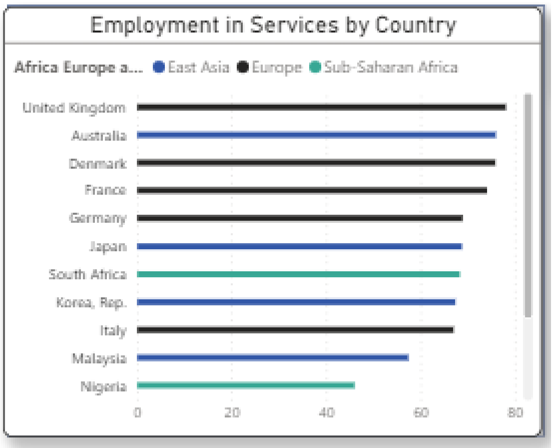


Figure 73. Visual 1

Visual 2: Does Urbanization Affect Employment Service?

This title of visual 2 in Figure 74 aims to captivate the audience by posing a question about the relationship between urbanization and employment in services. A combo chart is chosen to visually depict the correlation. The left scale represents the percentage of urbanization, while the right scale indicates employment in services. Categorical attribute, the country is on the x-axis, cont variable of wage and salaried workers is on the y-axis. Y-axis values are sorted from high to low. The visual demonstrates that countries experiencing higher urbanization also tend to have increased employment in services. While a strong correlation exists, it's essential to note that this relationship is not necessarily causal. The analysis spans countries with varying degrees of urbanization, offering nuanced insights into the dynamics of urbanization and employment in services. Countries such as the United Kingdom and France have high urbanization and high Employment showing a strong relationship while Kenya is the lowest according to the graph.

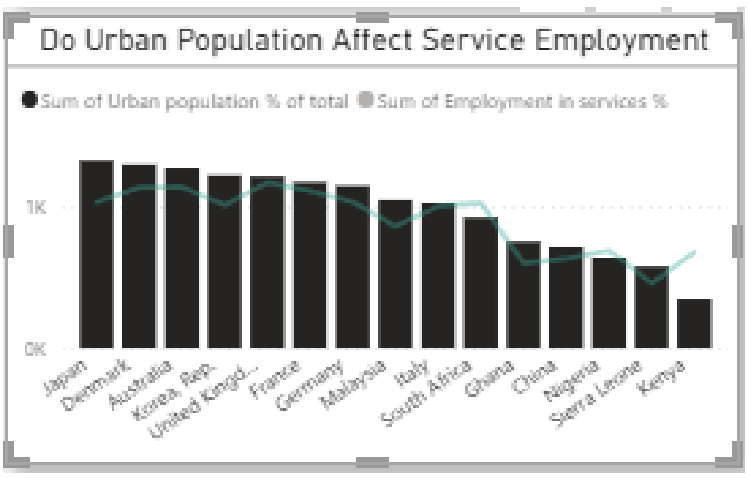


Figure 74. Visual 2

Visual 3: Year, Region and Country Slicer

The Year, Country, and Region Slicer is positioned directly below the title in Figure 75, allowing it to exhibit the specific year, country, or region to which the data pertains. When clicking on a particular year, it showcases data exclusively for that year and emphasizes the other visuals. Additionally, it provides the option to 'Select all,' offering average values for all the years collectively.

Figure 75. Slicer

Visual 4: Labour Force Participation.

The Count value was used to display the count value of labour force participation in visual 4, Figure 76, depending on the timeframe, region and year selected while relating it to other visuals in the dashboard. All visuals incorporate curved edges for a seamless appearance which offers visual appeal by reducing cognitive load and maintaining a smooth user reading experience during transitions between visuals as stated earlier in the overview of the dashboard

Figure 76. Count Visual

Visual 5: GDP Per Capital by Country and Region

This visual in Figure 77 employs a map to illustrate the influence of GDP on Employment in Services across different regions or countries. The map visual provides a geographical representation, allowing viewers to discern patterns and variations in Employment in Services concerning the respective GDP levels. Different shades or markers on the map may indicate the intensity or scale of this impact, providing a comprehensive overview of the relationship between GDP and Employment in Services on a global or regional scale.

Figure 77. Map

Visual 6. Do Wages Affect Employment Service?

This title is crafted to pique curiosity and engage the audience in exploring the intricate relationship between wages and employment in services. Employing a combo chart, similar to previous visuals, the left scale reflects wages, and the bottom of each bar depicts the countries. The visual in Figure 78 illustrates patterns suggesting that countries with higher wage levels often exhibit increased employment in services. It is crucial to emphasize that while a correlation is evident, causation cannot be assumed. The analysis spans a spectrum of countries, shedding light on the diverse impact of wages on employment in services. Notably, Italy stands out as an outlier, displaying average wage rates yet significantly low employment in service.

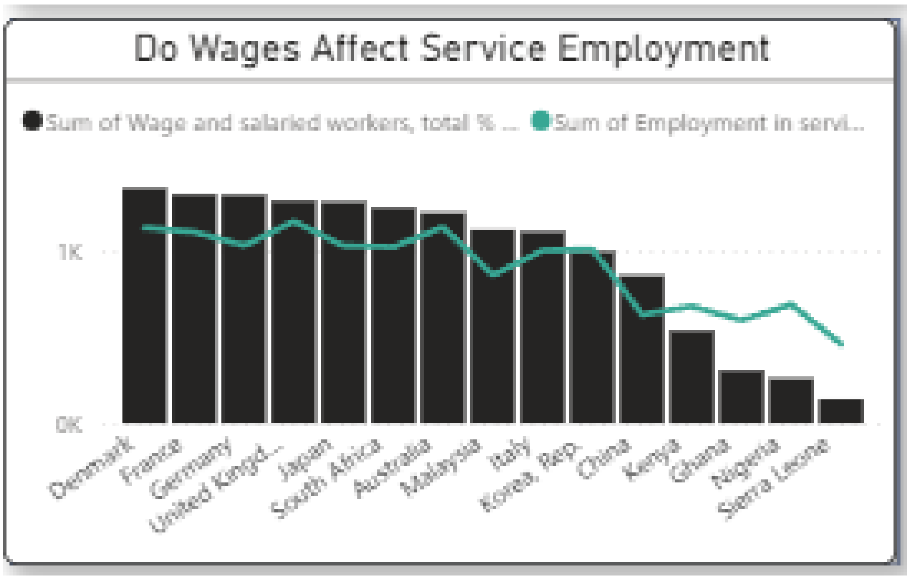


Figure 78. Visual 6

Visual 7: Forecasting Employment in Services

Within Power BI, the forecasting feature, facilitated by exponential smoothing, is a valuable tool. The chosen Employment in Service series, recorded annually without missing values, aligns seamlessly with this tool's requirements. Figure 79 depicts the Forecast Settings, wherein the units are converted to years, the forecast duration is expanded to 10 years, and a 95% confidence interval is specified. The seasonality is set to auto, enabling Power BI to identify its occurrence. It's crucial to emphasize that the default display of this visual showcases the average Employment in Service for all countries in both groups. However, the forecast projection for Employment in Service varies for each country. Leveraging the dashboard's interactivity, selecting any country in other visuals automatically adjusts the Forecast visual to display the Employment in Service forecast specifically for that chosen country.

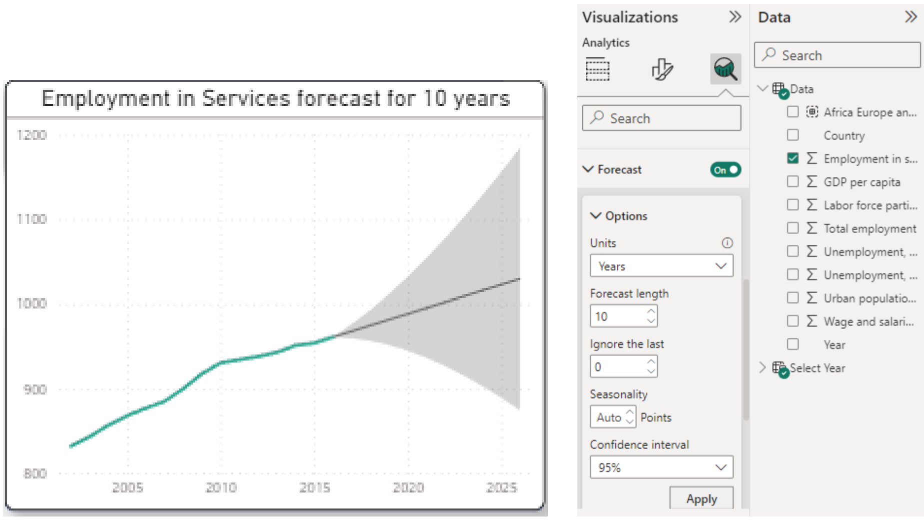


Figure 79. Forecast

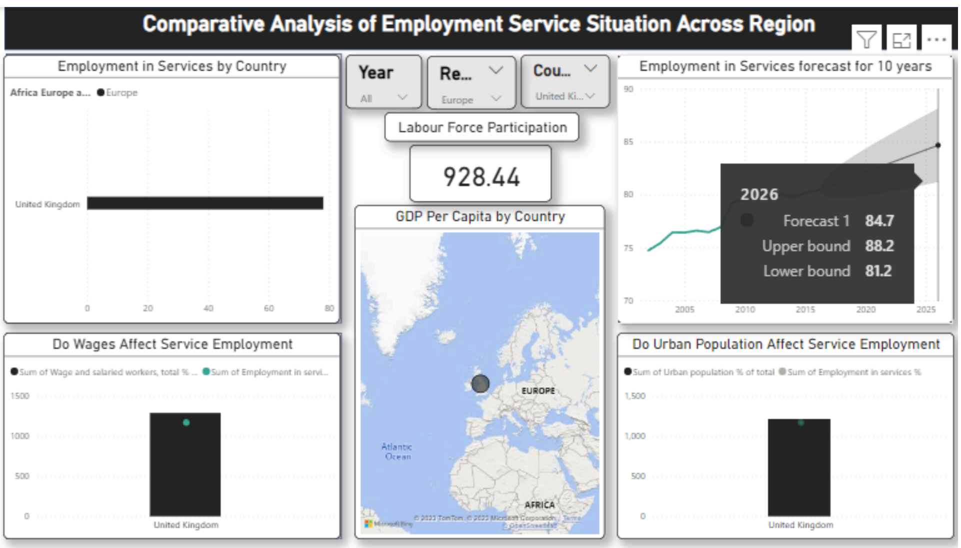


Figure 80. Forecast 1, Europe

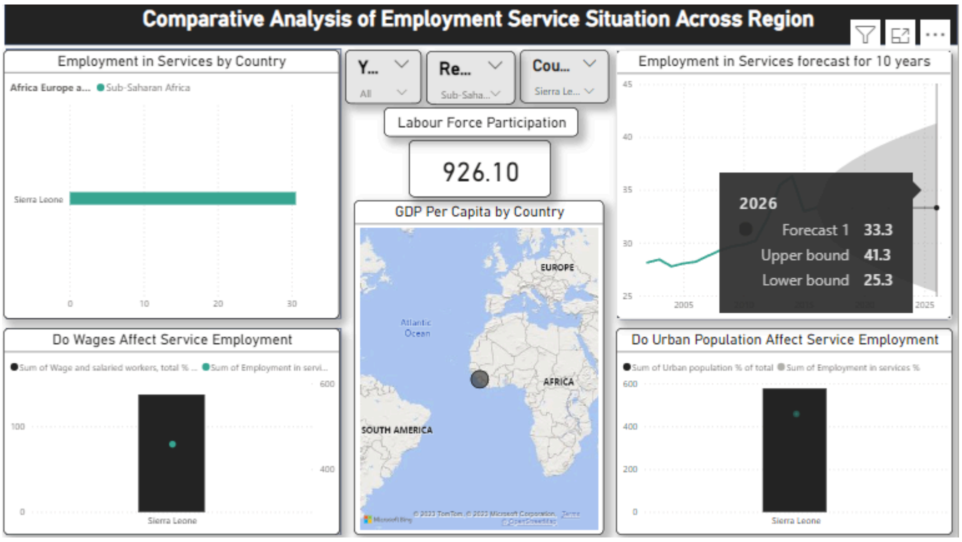


Figure 81. Forecast 2, Erria Leone

2.0 Discussion

The investigation into data workflows and the proposed dashboard design is a comprehensive process that involves understanding the characteristics of the data attributes, the iterative steps in designing the dashboard, and the thoughtful utilization of Power BI tools. The dataset selected consists of categorical attributes such as years and countries, with continuous numerical values representing various indicators. The primary target variable, employment in the service sector, is explored with other variables like wage and salaried workers, GDP per capita, urbanisation, and population. The dashboard design process involves meticulous steps, from attribute selection and data retrieval to cleaning and analysis. Power Query Editor is employed for data preprocessing, ensuring effective utilization and clarity. The layout planning is iterative, aligning with research goals, and the creation of insightful visuals using Power BI tools enhances the dashboard's informative capacity.

The discussion highlights the significance of data preprocessing, showcasing how data types, connections, decimal places, and categorisation are strategically handled. The dashboard's visual elements are carefully chosen to convey meaningful insights, with a focus on minimizing clutter and noise. Rounded corners and a minimalistic design are adopted for aesthetic appeal and user-friendly transitions between visuals. Each visual in the dashboard is explained, providing rationale and insights into the relationships depicted. Visual 1, for instance, reveals global economic trends in service sector employment, while Visual 2 explores the correlation between urbanization and employment in services. The inclusion of slicers and forecasting features adds interactivity and predictive capabilities to the dashboard. The overall design adheres to best practices, utilizing a colour scheme, minimalist elements, and carefully crafted titles to enhance user engagement. Rounded corners, as per UX Movement recommendations, contribute to a smooth user experience during transitions between visuals.

In summary, the project-based research demonstrates a systematic approach to designing a dashboard that not only visualizes complex relationships in the data but also considers user experience and engagement. The incorporation of Power BI tools, data preprocessing, and careful visual design contribute to the effectiveness of the dashboard in conveying meaningful insights from the dataset.

3.0 Conclusion

In conclusion, the research embarked on a comprehensive exploration of data workflows and proposed an intricately designed dashboard to fulfil its primary objectives. The investigation delved into understanding the characteristics of data attributes, emphasizing categorical and continuous variables, with a particular focus on employment in the service sector. The chosen dataset, sourced from Databank.org, underwent meticulous preprocessing using Power Query Editor, ensuring effective utilization and clarity for subsequent analyses.

The stepwise layout planning and creation of insightful visuals using Power BI tools were instrumental in achieving the research goals. The designed dashboard provides a holistic view of the relationships between employment in services and key variables such as wage and salaried workers, GDP per capita, urbanization, and population. The inclusion of slicers, combo charts, and forecasting features enhances the dashboard's interactivity and predictive capabilities.

The research emphasizes the importance of meaningful data preprocessing, thoughtful dashboard design, and adherence to best practices in visual representation. By adopting a minimalist design with rounded corners and a carefully chosen colour scheme, the dashboard ensures a smooth user experience, aligning with recommendations from UX Movement.

Overall, the designed dashboard serves as a powerful tool for exploring and communicating complex relationships within the dataset. The research successfully achieves its objectives, providing a valuable resource for understanding and interpreting the dynamics of employment in the service sector to various influential factors.

Part Four: References and Appendices

References

10 Best Practices for Building Effective Dashboards. (2022). https://www.tableau.com/learn/whitepapers/10-best-practices-building-effective-dashboards

Arévalo, S., & Devan, A. (2017). Unveiling storytelling and visualisation of data, *14th Student Colloquium at the University of Groningen, Groningen.* https://www.researchgate.net/publication/317236927_Unveiling_storytelling_and_visualization_of_data

Autor, D. H. (2019). Temporal Trends in Labor Market Dynamics: A Comprehensive Analysis, *Quarterly Journal of Economics,* (*34*)*(1)*.

Becker, G. S., & Huselid, M. A. (2006). Globalization and Employment Trends: A Comparative Analysis. *Journal of International Economics*, (*30*)*(2)*.

Blanchard, O. J., & Wolfers, J. (2000). Labor Market Dynamics Over the Business Cycle. *Journal of Economic Perspectives*, (*14*)(3).

Blinder, A. S. (2009). *Labor Market Dynamics in a Globalized Economy*, (115)(502), Economic Journal.

DataBank. (n.d.). The World Bank: https://databank.worldbank.org/metadataglossary/world-development-indicators/series

Few, S. (2004, March 20). Dashboard Confusion. *Intelltgent Enterprise.*

Few, S. (2010). *Dashboard Design for at-a-glance monitoring.* Perceptual Edge.

Fingleton, B., et al. (2012). Assessing Regional Economic Development Policies, *Journal of Regional Science.* (*28*)(4).

Glaeser, E. L., et al. (2009), Regional Economic Disparities: Causes and Consequences, *Journal of Economic Perspectives*, *24*(3).

Gupta, R., & Sharma, A. (2021), Regional Disparities in Employment: An Empirical Analysis. *Journal of Economic Geography, 21*(3).

Johnson, A., et al. (2017), Economic Development and Employment Trends: A Global Correlation Analysis. *International Journal of Economic Studies*, *(40)*(2).

Joost, de Winter & Gosling, Samuel & Potter, J.. (2016). *Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data.* Psychological Methods. 21. 273-290. 10.1037/met0000079.supp.

Jones, P., Smith, M., & Brown, T. (2018). Urbanization and Employment Patterns: A Comparative Study. *Journal of Urban Studies*. *45(2)*.

Holt-Winters (Time Series). (n.d.). Rapidminer. https://docs.rapidminer.com/latest/studio/operators/modeling/time_series/forecasting/holt-winters_trainer.html

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice (2nd ed.).* Melbourne, Australia: OTexts.com/fpp2. Retrieved December 9, 2022.

Hypothesis Testing. (n.d.). Retrieved 12 09, 2022, from Britannica. https://www.britannica.com/science/statistics/Hypothesis-testing

Kim, S., & Patel, R. (2018). GDP per Capita and Employment Dynamics: A Regression Analysis. World Economic Review, *45*(1).

Lee, Y. N., & Vivarelli, M. (2006). Correlating Socioeconomic Factors with Regional Employment: An Empirical Analysis, *Applied Economics, 15*(3).

Ma N, Shum WY, Han T and Cheong TS (2021) Global Inequality in Service Sector Development: Trend and Convergence. *Front. Psychol.* 12:792950. doi: 10.3389/fpsyg.2021.792950

Madhavan, A. (2019). *Causation vs Correlation. Amplitude.* https://amplitude.com/blog/causation-correlation

Martin, R., & Sunley, P. (2015). Understanding Drivers of Regional Employment Disparities. *Economic Geography,* 40*(1)*.

Morgan, K. (2020, April 5). *How People Read Online: New and Old Findings.* https://www.nngroup.com/reports/how-people-read-web-eyetracking-evidence/

Oosthuizen, M. (n.d.). *Key influencers - Power BI Al Visuals* https://databear.com/key-influencers-power-bi-ai-visuals

R & Data Mining (n.d.). STHDA http://www.sthda.com/english/wiki/correlation-test-between-two-variables-in-r

R Programming Language. (2022). https://www.geeksforgeeks.org/apply-lapply-sapply-and-tapply-in-r/

Rios-Salas, V., & Trupp, A. (2014). Socioeconomic Factors and Regional Employment Patterns: A Comparative Study. *Regional Studies, 22*(4).

Rodriguez, C., & Chen, L. (2015). Urbanization and Employment: A Descriptive Analysis, *Journal of Urban Economics, 28*(3).

Rodriguez-Pose, A., & Hardy, D. (2015). Regional Economic Development Strategies and Employment Disparities. *Regional Studies*, *18*(2).

Salas-Velasco, M., & López-Bazo, E. (2020). Exploring Regional Wage Differentials: Evidence from a Multilevel Approach, *Regional Studies,* (*54*)(7).

Smith, J., & Brown, M. (2019), "Education and Employment: A Hypothesis Testing Approach, *Journal of Labor Economics Research*, (*27*)(4).

Smith, J., Williams, K., & Anderson, R. (2019), *Understanding the Social Impact of Regional Job Disparities, *Journal of Regional Development,* (*35*)(4).

Soetewey. A. (2020, July 5). *Stats and R.* https://statsandr.com/blog/wilcoxon-test-in-r-how-to-compare-2-groups-under-the-non-normality-assumption/

Skewness and Kurtosis in Statistics. (n.d.). r-blogger.com https://www.r-bloggers.com/2020/11/skewness-and-kurtosis-in-statistics/

Statistical Hypothesis Testing. (n.d.). *In Wikipedia*. https://en.wikipedia.org/wiki/Statistical_hypothesis_testing

Steele, R., et al. (2011). Beautiful Visualization: Looking at Data through the Eyes of Experts. O'Reilly Media.

The Principles of Visual Design for Dashboards. (2021). https://adniasolutions.com/dashboard-design-principles/the-principles-of-visual-design-for-dashboards/

Time Series the basics. (n.d.). Australia Bureau of Statistics. https://www.abs.gov.au/websitedbs/d3310114.nsf/home/time+series+analysis:+the+basIcs

Tseng, A. (2011, August 17). *Why Rounded Corners Are Easier on the Eyes*. https://uxmovement.com/thinking/why-rounded-corners-are-easier-on-the-eyes/

Visualization types in Power BI. (2022, 02 09). https://learn.microsoft.com/en-us/power-bi/visuals/power-bi-visualization-types-for-reports-and-q-and-a#combo-charts

Wexler, S., Shaffer, J., & Cotgreave, A. (2017). *The Big Book of Dashboards. Hoboken, New Jersey: John Wiley &Sons, Inc.*

Wooldridge, J. (2002). *Introductory Econometrics: A Modem Approach.*

World development indicators. (n.d.). World Bank. https://databank.worldbank.org/source/world-development-indicators