

Web_Scraping

Thrishin

2023-10-16

Setting the R environment

Installing the packages

Notes: The code installs three essential R packages. “tidyverse” is a versatile data manipulation and visualization library. “rvest” is used for web scraping. “rmarkdown” is for creating dynamic documents with R code

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
install.packages('rvest')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
install.packages("rmarkdown")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

Loading the installed packages

Notes: The provided code snippet loads two R packages. “tidyverse” offers data manipulation and visualization tools, while “rvest” is used for web scraping. These libraries are now accessible for analysis.

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.3      v readr      2.1.4  
## v forcats    1.0.0      v stringr   1.5.0  
## v ggplot2    3.4.4      v tibble    3.2.1  
## v lubridate  1.9.3      v tidyr     1.3.0  
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(rvest)
```

```
##
```

```
## Attaching package: 'rvest'
```

```
##
```

```
## The following object is masked from 'package:readr':
```

```
##  
##      guess_encoding
```

STEP 1: Web Scraping

Web scraping the World Happiness report from Wikipedia

Notes: This code snippet retrieves data from a Wikipedia page on the “World Happiness Report” using the specified URL. It utilizes the “rvest” package to read and parse the HTML content of the web page. The data of interest is extracted from a specific HTML table using an XPath selector and converted into a data frame for further analysis.

```
url <- "https://en.wikipedia.org/wiki/World_Happiness_Report#2017_report"  
  
page <- read_html(url)  
  
table_data <- page %>%  
  html_element(xpath = '//*[@id="mw-content-text"]/div[1]/div[18]/table/tbody/tr[2]/td/table') %>%  
  html_table()
```

Loading another happiness report table

Notes: This commented-out code section was intended to load another happiness report table from a different location on the same web page. It uses the “rvest” package to scrape and convert data from a specific HTML table using XPath.

```
#table_data1 <- page %>%  
#   html_element(xpath = '//*[@id="mw-content-text"]/div[1]/div[15]/table/tbody/tr[2]/td/table') %>%  
#   html_table()
```

STEP 2 : Cleaning the web scraped data

Loading the required library

Notes: This code loads the “dplyr” library, enabling efficient data manipulation in R. “dplyr” provides a range of functions for data filtering, transformation, and summarization, enhancing data analysis capabilities.

```
library(dplyr)
```

Convert column names to lowercase for consistency

Notes: This code converts all column names in the “table_data” data frame to lowercase letters for uniformity, simplifying subsequent data manipulation and analysis processes.

```
table_data <- table_data %>%  
  rename_all(tolower)
```

To view the table where the change is made

Notes: This code, displays the “table_data” data frame in a viewer or plot window, allowing users to inspect the changes made to column names, ensuring data consistency.

```
view(table_data)
```

Replace spaces with underscores in column names

Notes: This code substitutes spaces with underscores in all column names of the “table_data” data frame. After executing this transformation, the viewer displays the updated dataset, improving data consistency for

analysis and enabling compatibility with R's conventions for naming variables.

```
table_data <- table_data %>%  
  rename_with(~ gsub(" ", "_", .), everything())  
view(table_data)
```

Remove rows with missing values (if any)

Notes: This code eliminates rows with missing values in the “table_data” data frame using the “na.omit” function. It ensures a cleaner dataset for analysis by removing incomplete or problematic records.

```
table_data <- table_data %>%  
  na.omit()
```

```
view(table_data)
```

Replace “N/A” with 0 in all columns

Notes: This code replaces all occurrences of “N/A” with “0” in all columns of the “table_data” data frame, ensuring consistent data for analysis by substituting missing or non-numeric values with zeros.

```
table_data <- data.frame(lapply(table_data, function(x) gsub("N/A", "0", x)))
```

Convert “score” to numeric, replacing non-numeric values with NA

Notes: This code converts the “score” column in the “table_data” data frame to numeric, replacing non-numeric values with NA. It filters out rows with non-numeric score values, ensuring data integrity for analysis. The viewer displays the updated dataset.

```
unique(table_data$score)
```

```
## [1] "7.537" "7.522" "7.504" "7.494" "7.469" "7.377"  
## [7] "7.316" "7.314" "7.284" "7.213" "7.079" "7.006"  
## [13] "6.993" "6.977" "6.951" "6.891" "6.863" "6.714"  
## [19] "6.652" "6.648" "6.635" "6.609" "6.599" "6.578"  
## [25] "6.572" "6.527" "6.454" "6.452" "6.442" "6.424"  
## [31] "6.422" "6.403" "6.375" "6.357" "6.344" "6.168"  
## [37] "6.105" "6.098" "6.087" "6.084" "6.080" "6.071"  
## [43] "6.008" "6.003" "5.973" "5.971" "5.964" "5.963"  
## [49] "5.956" "5.920" "5.902" "5.872" "5.850" "5.838"  
## [55] "5.825" "5.823" "5.822" "5.819" "5.810" "5.758"  
## [61] "5.715" "5.629" "5.621" "5.611" "5.569" "5.525"  
## [67] "5.500" "5.493" "5.472" "5.430" "5.395" "5.336"  
## [73] "5.324" "5.311" "5.305[b] " "5.293" "5.279" "5.273"  
## [79] "5.269" "5.262" "5.250" "5.237" "5.235" "7.342"  
## [85] "5.230" "5.227" "5.225" "5.195" "5.182" "5.181"  
## [91] "5.175" "5.151" "5.074" "5.041" "5.011" "5.004"  
## [97] "4.962" "4.955" "4.829" "4.805" "4.775" "4.735"  
## [103] "4.714" "4.709" "4.695" "4.692" "4.644" "4.608"  
## [109] "4.574" "4.553" "4.550" "4.545" "4.535" "4.514"  
## [115] "4.497" "4.465" "4.460" "4.440" "4.376" "4.315"  
## [121] "4.292" "4.291" "4.286" "4.280" "4.190" "4.180"  
## [127] "4.168" "4.139" "4.120" "4.096" "4.081" "4.032"  
## [133] "4.028" "3.970" "3.936" "3.875" "3.808" "3.795"  
## [139] "3.794" "3.766" "3.657" "3.644" "3.603" "3.593"  
## [145] "3.591" "3.533" "3.507" "3.495" "3.471" "3.462"  
## [151] "3.349" "2.905" "2.693"
```

```
table_data$score <- as.numeric(table_data$score)

## Warning: NAs introduced by coercion

table_data <- table_data[!is.na(as.numeric(table_data$score)), ]

view(table_data)
```

Summary statistics

Notes: This code computes summary statistics for specific columns in the “table_data” data frame. It calculates the mean score, minimum GDP, and maximum generosity values. The results are displayed in a viewer for quick data insight.

```
summary_stats <- table_data %>%
  summarise(
    Mean_Score = mean(score),
    Min_GDP = min(gdp_per_capita),
    Max_Generosity = max(generosity)
  )

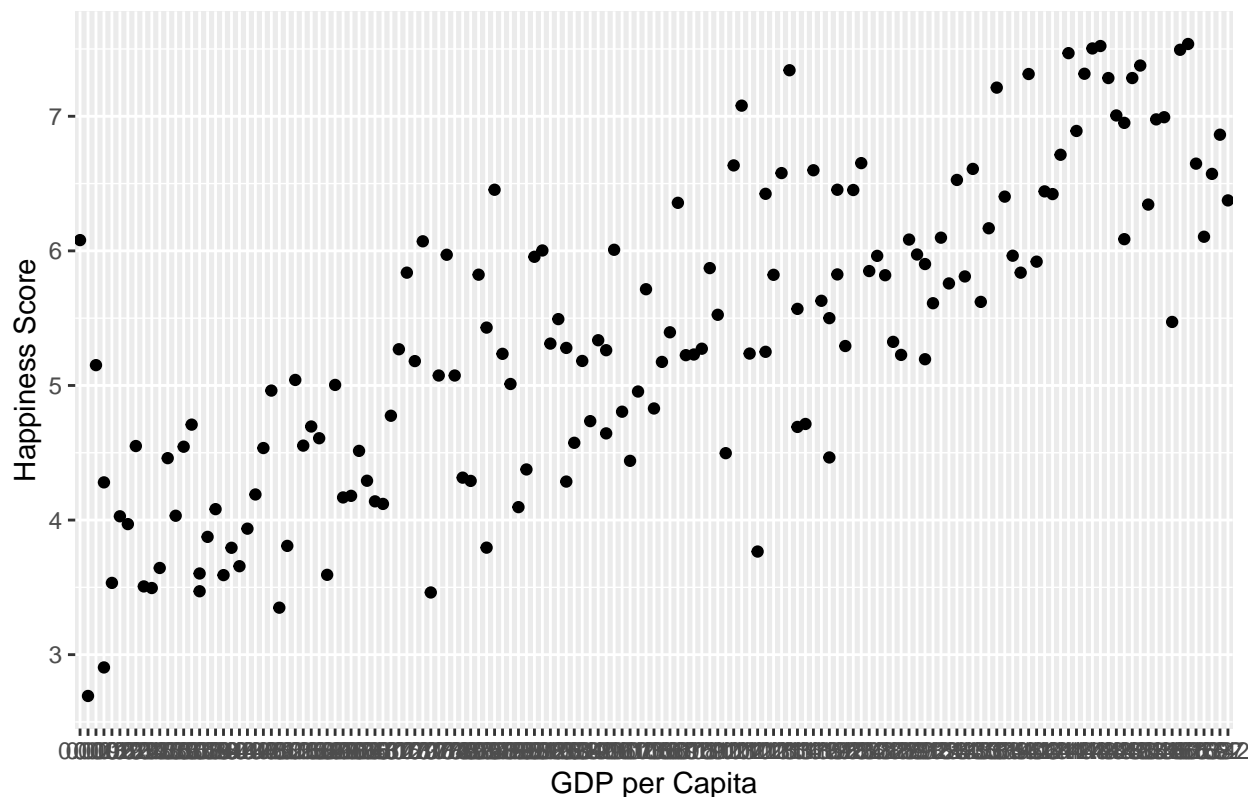
view(summary_stats)
```

Scatter Plot of GDP per Capita vs. Happiness Score

Notes: This code generates a scatter plot using “ggplot2,” with GDP per capita on the x-axis and happiness score on the y-axis. The plot is given a title and labeled axes, providing a visual representation of their relationship.

```
ggplot(table_data, aes(x = gdp_per_capita, y = score)) +
  geom_point() +
  labs(title = "Scatter Plot of GDP per Capita vs. Happiness Score", x = "GDP per Capita", y = "Happiness Score")
```

Scatter Plot of GDP per Capita vs. Happiness Score



STEP 4: Exploring Relationship

Explore relationships between variables. You can create a correlation matrix, for example.

Notes: This code calculates a correlation matrix, exploring relationships between specified variables in the “table_data” data frame. The matrix is printed, providing insights into the associations among variables like score, GDP per capita, social support, and more.

```
# correlation_matrix <- cor(table_data %>% select(score, gdp_per_capita, social_support, healthy_life_e
# print(correlation_matrix)
```

STEP 5

Hypothesis Testing if there is a significant difference in scores between two groups (e.g., high GDP vs. low GDP countries)

Notes: This code conducts a t-test to determine if a significant difference exists in scores between two groups, such as countries with high GDP per capita (greater than or equal to 1) and low GDP per capita (less than 1). It assesses score disparities between these groups.

```
t.test(table_data[table_data$gdp_per_capita >= 1, "score"], table_data[table_data$gdp_per_capita < 1, "score"],
       ##
       ## Welch Two Sample t-test
       ##
       ## data: table_data[table_data$gdp_per_capita >= 1, "score"] and table_data[table_data$gdp_per_capita < 1, "score"]
       ## t = 11.464, df = 152.23, p-value < 2.2e-16
       ## alternative hypothesis: true difference in means is not equal to 0
       ## 95 percent confidence interval:
```

```
## 1.277803 1.809923
## sample estimates:
## mean of x mean of y
## 6.084738 4.540875
```

Modeling happiness scores based on GDP per capita

Notes: This code creates a linear regression model to analyze the relationship between happiness scores and GDP per capita in the “table_data” dataset. The “summary” function provides insights into the model’s statistical significance and coefficients.

```
linear_model <- lm(score ~ gdp_per_capita, data = table_data)
summary(linear_model)
```

```
##
## Call:
## lm(formula = score ~ gdp_per_capita, data = table_data)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-0.8175	0.0000	0.0000	0.0000	0.8175

```
##
## Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)	
## (Intercept)	6.0800	0.7010	8.674	3e-06	***
## gdp_per_capita0.000	-3.3870	0.9913	-3.417	0.00576	**
## gdp_per_capita0.023	-0.9290	0.9913	-0.937	0.36881	
## gdp_per_capita0.092	-2.4875	0.8585	-2.897	0.01451	*
## gdp_per_capita0.119	-2.5470	0.9913	-2.569	0.02608	*
## gdp_per_capita0.162	-2.0520	0.9913	-2.070	0.06278	.
## gdp_per_capita0.233	-2.1100	0.9913	-2.128	0.05673	.
## gdp_per_capita0.234	-1.5300	0.9913	-1.543	0.15100	
## gdp_per_capita0.245	-2.5730	0.9913	-2.596	0.02489	*
## gdp_per_capita0.305	-2.5850	0.9913	-2.608	0.02436	*
## gdp_per_capita0.306	-2.4360	0.9913	-2.457	0.03183	*
## gdp_per_capita0.339	-1.6200	0.9913	-1.634	0.13049	
## gdp_per_capita0.350	-2.0480	0.9913	-2.066	0.06322	.
## gdp_per_capita0.367	-1.5350	0.9913	-1.548	0.14979	
## gdp_per_capita0.368	-1.3710	0.9913	-1.383	0.19409	
## gdp_per_capita0.369	-2.5430	0.8585	-2.962	0.01293	*
## gdp_per_capita0.376	-2.2050	0.9913	-2.224	0.04800	*
## gdp_per_capita0.381	-1.9990	0.9913	-2.017	0.06883	.
## gdp_per_capita0.397	-2.4890	0.9913	-2.511	0.02895	*
## gdp_per_capita0.401	-2.2860	0.9913	-2.306	0.04159	*
## gdp_per_capita0.431	-2.4230	0.9913	-2.444	0.03258	*
## gdp_per_capita0.438	-2.1440	0.9913	-2.163	0.05345	.
## gdp_per_capita0.476	-1.8900	0.9913	-1.907	0.08302	.
## gdp_per_capita0.479	-1.5450	0.9913	-1.559	0.14740	
## gdp_per_capita0.480	-1.1180	0.9913	-1.128	0.28341	
## gdp_per_capita0.511	-2.7310	0.9913	-2.755	0.01873	*
## gdp_per_capita0.521	-2.2720	0.9913	-2.292	0.04264	*
## gdp_per_capita0.525	-1.0390	0.9913	-1.048	0.31708	
## gdp_per_capita0.560	-1.5270	0.9913	-1.540	0.15173	
## gdp_per_capita0.564	-1.3850	0.9913	-1.397	0.18992	
## gdp_per_capita0.587	-1.4720	0.9913	-1.485	0.16566	

## gdp_per_capita0.592	-2.4870	0.9913	-2.509	0.02905 *
## gdp_per_capita0.596	-1.0760	0.9913	-1.085	0.30095
## gdp_per_capita0.602	-1.9120	0.9913	-1.929	0.07996 .
## gdp_per_capita0.603	-1.9000	0.9913	-1.917	0.08162 .
## gdp_per_capita0.636	-1.5660	0.9913	-1.580	0.14248
## gdp_per_capita0.648	-1.7880	0.9913	-1.804	0.09871 .
## gdp_per_capita0.660	-1.9410	0.9913	-1.958	0.07607 .
## gdp_per_capita0.667	-1.9600	0.9913	-1.977	0.07363 .
## gdp_per_capita0.716	-1.3050	0.9913	-1.316	0.21480
## gdp_per_capita0.727	-0.8110	0.9913	-0.818	0.43066
## gdp_per_capita0.729	-0.2420	0.9913	-0.244	0.81163
## gdp_per_capita0.731	-0.8990	0.9913	-0.907	0.38391
## gdp_per_capita0.737	-0.0090	0.9913	-0.009	0.99292
## gdp_per_capita0.777	-2.6180	0.9913	-2.641	0.02296 *
## gdp_per_capita0.784	-1.0060	0.9913	-1.015	0.33200
## gdp_per_capita0.786	-0.1090	0.9913	-0.110	0.91443
## gdp_per_capita0.789	-1.0060	0.9913	-1.015	0.33200
## gdp_per_capita0.792	-1.7650	0.9913	-1.780	0.10260
## gdp_per_capita0.809	-1.7890	0.9913	-1.805	0.09855 .
## gdp_per_capita0.834	-0.2570	0.9913	-0.259	0.80023
## gdp_per_capita0.858	-1.4675	0.8585	-1.709	0.11541
## gdp_per_capita0.872	0.3740	0.9913	0.377	0.71315
## gdp_per_capita0.878	-0.8450	0.9913	-0.852	0.41217
## gdp_per_capita0.885	-1.0690	0.9913	-1.078	0.30395
## gdp_per_capita0.895	-1.9840	0.9913	-2.001	0.07064 .
## gdp_per_capita0.901	-1.7040	0.9913	-1.719	0.11361
## gdp_per_capita0.908	-0.1240	0.9913	-0.125	0.90271
## gdp_per_capita0.910	-0.0770	0.9913	-0.078	0.93948
## gdp_per_capita0.926	-0.7690	0.9913	-0.776	0.45426
## gdp_per_capita0.933	-0.5870	0.9913	-0.592	0.56573
## gdp_per_capita0.951	-1.2975	0.8585	-1.511	0.15888
## gdp_per_capita0.964	-1.5060	0.9913	-1.519	0.15692
## gdp_per_capita0.982	-0.8980	0.9913	-0.906	0.38442
## gdp_per_capita0.990	-1.3450	0.9913	-1.357	0.20204
## gdp_per_capita0.991	-0.7440	0.9913	-0.751	0.46870
## gdp_per_capita0.996	-1.1270	0.8585	-1.313	0.21600
## gdp_per_capita1.001	-0.0720	0.9913	-0.073	0.94340
## gdp_per_capita1.007	-1.2750	0.9913	-1.286	0.22480
## gdp_per_capita1.010	-1.6400	0.9913	-1.654	0.12628
## gdp_per_capita1.027	-1.1250	0.9913	-1.135	0.28056
## gdp_per_capita1.035	-0.3650	0.9913	-0.368	0.71971
## gdp_per_capita1.055	-1.2510	0.9913	-1.262	0.23307
## gdp_per_capita1.065	-0.9050	0.9913	-0.913	0.38086
## gdp_per_capita1.069	-0.6850	0.9913	-0.691	0.50389
## gdp_per_capita1.071	0.2770	0.9913	0.279	0.78510
## gdp_per_capita1.075	-0.8550	0.9913	-0.862	0.40683
## gdp_per_capita1.079	-0.8500	0.9913	-0.857	0.40949
## gdp_per_capita1.081	-0.8070	0.9913	-0.814	0.43288
## gdp_per_capita1.092	-0.2080	0.9913	-0.210	0.83764
## gdp_per_capita1.102	-0.5550	0.9913	-0.560	0.58679
## gdp_per_capita1.103	-1.5830	0.9913	-1.597	0.13860
## gdp_per_capita1.107	0.5550	0.9913	0.560	0.58679
## gdp_per_capita1.110	0.9990	0.9913	1.008	0.33523
## gdp_per_capita1.121	-0.8430	0.9913	-0.850	0.41324

## gdp_per_capita1.122	-2.3140	0.9913	-2.334	0.03957 *
## gdp_per_capita1.128	-0.2430	0.8585	-0.283	0.78239
## gdp_per_capita1.131	-0.2580	0.9913	-0.260	0.79947
## gdp_per_capita1.153	0.4980	0.9913	0.502	0.62532
## gdp_per_capita1.154	1.2620	0.9913	1.273	0.22925
## gdp_per_capita1.157	-0.9495	0.8585	-1.106	0.29233
## gdp_per_capita1.161	-1.3660	0.9913	-1.378	0.19560
## gdp_per_capita1.185	0.5190	0.9913	0.524	0.61098
## gdp_per_capita1.189	-0.4510	0.9913	-0.455	0.65799
## gdp_per_capita1.198	-1.0975	0.8585	-1.278	0.22743
## gdp_per_capita1.218	0.0595	0.8585	0.069	0.94599
## gdp_per_capita1.223	-0.7870	0.9913	-0.794	0.44405
## gdp_per_capita1.234	0.3720	0.9913	0.375	0.71460
## gdp_per_capita1.253	0.5720	0.9913	0.577	0.57555
## gdp_per_capita1.261	-0.2300	0.9913	-0.232	0.82079
## gdp_per_capita1.282	-0.1170	0.9913	-0.118	0.90818
## gdp_per_capita1.285	-0.2610	0.9913	-0.263	0.79720
## gdp_per_capita1.286	-0.7560	0.9913	-0.763	0.46173
## gdp_per_capita1.289	-0.8530	0.9913	-0.860	0.40789
## gdp_per_capita1.291	0.0040	0.9913	0.004	0.99685
## gdp_per_capita1.292	-0.1070	0.9913	-0.108	0.91599
## gdp_per_capita1.315	-0.5315	0.8585	-0.619	0.54846
## gdp_per_capita1.321	-0.4690	0.9913	-0.473	0.64539
## gdp_per_capita1.325	0.0180	0.9913	0.018	0.98584
## gdp_per_capita1.341	-0.3220	0.9913	-0.325	0.75141
## gdp_per_capita1.343	0.4470	0.9913	0.451	0.66081
## gdp_per_capita1.347	-0.2700	0.9913	-0.272	0.79039
## gdp_per_capita1.353	0.5290	0.9913	0.534	0.60421
## gdp_per_capita1.356	-0.4590	0.9913	-0.463	0.65238
## gdp_per_capita1.361	0.0880	0.9913	0.089	0.93086
## gdp_per_capita1.375	1.1330	0.9913	1.143	0.27734
## gdp_per_capita1.384	0.3230	0.9913	0.326	0.75067
## gdp_per_capita1.395	-0.1160	0.9913	-0.117	0.90896
## gdp_per_capita1.402	-0.2420	0.9913	-0.244	0.81163
## gdp_per_capita1.406	1.2340	0.9913	1.245	0.23907
## gdp_per_capita1.417	-0.1600	0.9913	-0.161	0.87470
## gdp_per_capita1.431	0.3620	0.9913	0.365	0.72191
## gdp_per_capita1.434	0.3420	0.9913	0.345	0.73660
## gdp_per_capita1.442	0.6340	0.9913	0.640	0.53556
## gdp_per_capita1.444	1.3890	0.9913	1.401	0.18874
## gdp_per_capita1.464	0.8110	0.9913	0.818	0.43066
## gdp_per_capita1.479	1.2360	0.9913	1.247	0.23836
## gdp_per_capita1.481	1.4240	0.9913	1.436	0.17869
## gdp_per_capita1.482	1.4420	0.9913	1.455	0.17371
## gdp_per_capita1.484	1.2040	0.9913	1.215	0.24997
## gdp_per_capita1.487	0.9260	0.9913	0.934	0.37030
## gdp_per_capita1.488	0.4390	0.8585	0.511	0.61921
## gdp_per_capita1.494	1.2040	0.9913	1.215	0.24997
## gdp_per_capita1.504	1.2970	0.9913	1.308	0.21743
## gdp_per_capita1.531	0.2640	0.9913	0.266	0.79493
## gdp_per_capita1.536	0.8970	0.9913	0.905	0.38493
## gdp_per_capita1.546	0.9130	0.9913	0.921	0.37681
## gdp_per_capita1.552	-0.6080	0.9913	-0.613	0.55213
## gdp_per_capita1.565	1.4140	0.9913	1.426	0.18152


```
## gdp_per_capita1.616    1.4570    0.9913    1.470    0.16964
## gdp_per_capita1.626    0.5680    0.9913    0.573    0.57819
## gdp_per_capita1.633    0.0250    0.9913    0.025    0.98033
## gdp_per_capita1.692    0.4920    0.9913    0.496    0.62945
## gdp_per_capita1.742    0.7830    0.9913    0.790    0.44630
## gdp_per_capita1.871    0.2950    0.9913    0.298    0.77157
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.701 on 11 degrees of freedom
## Multiple R-squared:  0.9732, Adjusted R-squared:  0.622
## F-statistic: 2.771 on 144 and 11 DF,  p-value: 0.03117
```

Linear Regression with Multiple Predictors (gdp_per_capita, social_support, healthy_life_expectancy)

Notes: This code fits a linear regression model with multiple predictors, including GDP per capita, social support, and healthy life expectancy, to understand their collective impact on happiness scores in the “table_data” dataset. The “summary” function presents statistical information about the model.

```
model <- lm(score ~ gdp_per_capita + social_support + healthy_life_expectancy, data = table_data)
summary(model)
```

```
##
## Call:
## lm(formula = score ~ gdp_per_capita + social_support + healthy_life_expectancy,
##     data = table_data)
##
## Residuals:
## ALL 156 residuals are 0: no residual degrees of freedom!
##
## Coefficients: (269 not defined because of singularities)
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.080         NaN      NaN    NaN
## gdp_per_capita0.000    -3.387         NaN      NaN    NaN
## gdp_per_capita0.023    -0.929         NaN      NaN    NaN
## gdp_per_capita0.092    -1.800         NaN      NaN    NaN
## gdp_per_capita0.119    -2.547         NaN      NaN    NaN
## gdp_per_capita0.162    -2.052         NaN      NaN    NaN
## gdp_per_capita0.233    -2.110         NaN      NaN    NaN
## gdp_per_capita0.234    -1.530         NaN      NaN    NaN
## gdp_per_capita0.245    -2.573         NaN      NaN    NaN
## gdp_per_capita0.305    -2.585         NaN      NaN    NaN
## gdp_per_capita0.306    -2.436         NaN      NaN    NaN
## gdp_per_capita0.339    -1.620         NaN      NaN    NaN
## gdp_per_capita0.350    -2.048         NaN      NaN    NaN
## gdp_per_capita0.367    -1.535         NaN      NaN    NaN
## gdp_per_capita0.368    -1.371         NaN      NaN    NaN
## gdp_per_capita0.369    -2.609         NaN      NaN    NaN
## gdp_per_capita0.376    -2.205         NaN      NaN    NaN
## gdp_per_capita0.381    -1.999         NaN      NaN    NaN
## gdp_per_capita0.397    -2.489         NaN      NaN    NaN
## gdp_per_capita0.401    -2.286         NaN      NaN    NaN
## gdp_per_capita0.431    -2.423         NaN      NaN    NaN
## gdp_per_capita0.438    -2.144         NaN      NaN    NaN
## gdp_per_capita0.476    -1.890         NaN      NaN    NaN
```

## gdp_per_capita0.479	-1.545	NaN	NaN	NaN
## gdp_per_capita0.480	-1.118	NaN	NaN	NaN
## gdp_per_capita0.511	-2.731	NaN	NaN	NaN
## gdp_per_capita0.521	-2.272	NaN	NaN	NaN
## gdp_per_capita0.525	-1.039	NaN	NaN	NaN
## gdp_per_capita0.560	-1.527	NaN	NaN	NaN
## gdp_per_capita0.564	-1.385	NaN	NaN	NaN
## gdp_per_capita0.587	-1.472	NaN	NaN	NaN
## gdp_per_capita0.592	-2.487	NaN	NaN	NaN
## gdp_per_capita0.596	-1.076	NaN	NaN	NaN
## gdp_per_capita0.602	-1.912	NaN	NaN	NaN
## gdp_per_capita0.603	-1.900	NaN	NaN	NaN
## gdp_per_capita0.636	-1.566	NaN	NaN	NaN
## gdp_per_capita0.648	-1.788	NaN	NaN	NaN
## gdp_per_capita0.660	-1.941	NaN	NaN	NaN
## gdp_per_capita0.667	-1.960	NaN	NaN	NaN
## gdp_per_capita0.716	-0.270	NaN	NaN	NaN
## gdp_per_capita0.727	-0.811	NaN	NaN	NaN
## gdp_per_capita0.729	-0.242	NaN	NaN	NaN
## gdp_per_capita0.731	-0.899	NaN	NaN	NaN
## gdp_per_capita0.737	-0.009	NaN	NaN	NaN
## gdp_per_capita0.777	-2.618	NaN	NaN	NaN
## gdp_per_capita0.784	-1.006	NaN	NaN	NaN
## gdp_per_capita0.786	-0.109	NaN	NaN	NaN
## gdp_per_capita0.789	-1.006	NaN	NaN	NaN
## gdp_per_capita0.792	-1.765	NaN	NaN	NaN
## gdp_per_capita0.809	-1.789	NaN	NaN	NaN
## gdp_per_capita0.834	-0.257	NaN	NaN	NaN
## gdp_per_capita0.858	-0.650	NaN	NaN	NaN
## gdp_per_capita0.872	0.374	NaN	NaN	NaN
## gdp_per_capita0.878	-0.845	NaN	NaN	NaN
## gdp_per_capita0.885	-1.069	NaN	NaN	NaN
## gdp_per_capita0.895	-1.984	NaN	NaN	NaN
## gdp_per_capita0.901	-1.704	NaN	NaN	NaN
## gdp_per_capita0.908	-0.124	NaN	NaN	NaN
## gdp_per_capita0.910	-0.077	NaN	NaN	NaN
## gdp_per_capita0.926	-0.769	NaN	NaN	NaN
## gdp_per_capita0.933	-0.587	NaN	NaN	NaN
## gdp_per_capita0.951	-0.801	NaN	NaN	NaN
## gdp_per_capita0.964	-1.506	NaN	NaN	NaN
## gdp_per_capita0.982	-0.898	NaN	NaN	NaN
## gdp_per_capita0.990	-1.345	NaN	NaN	NaN
## gdp_per_capita0.991	-0.744	NaN	NaN	NaN
## gdp_per_capita0.996	-0.818	NaN	NaN	NaN
## gdp_per_capita1.001	-0.072	NaN	NaN	NaN
## gdp_per_capita1.007	-1.275	NaN	NaN	NaN
## gdp_per_capita1.010	-1.640	NaN	NaN	NaN
## gdp_per_capita1.027	-1.125	NaN	NaN	NaN
## gdp_per_capita1.035	-0.365	NaN	NaN	NaN
## gdp_per_capita1.055	-1.251	NaN	NaN	NaN
## gdp_per_capita1.065	-0.905	NaN	NaN	NaN
## gdp_per_capita1.069	-0.685	NaN	NaN	NaN
## gdp_per_capita1.071	0.277	NaN	NaN	NaN
## gdp_per_capita1.075	-0.855	NaN	NaN	NaN

## gdp_per_capita1.079	-0.850	NaN	NaN	NaN
## gdp_per_capita1.081	-0.807	NaN	NaN	NaN
## gdp_per_capita1.092	-0.208	NaN	NaN	NaN
## gdp_per_capita1.102	-0.555	NaN	NaN	NaN
## gdp_per_capita1.103	-1.583	NaN	NaN	NaN
## gdp_per_capita1.107	0.555	NaN	NaN	NaN
## gdp_per_capita1.110	0.999	NaN	NaN	NaN
## gdp_per_capita1.121	-0.843	NaN	NaN	NaN
## gdp_per_capita1.122	-2.314	NaN	NaN	NaN
## gdp_per_capita1.128	-0.830	NaN	NaN	NaN
## gdp_per_capita1.131	-0.258	NaN	NaN	NaN
## gdp_per_capita1.153	0.498	NaN	NaN	NaN
## gdp_per_capita1.154	1.262	NaN	NaN	NaN
## gdp_per_capita1.157	-0.511	NaN	NaN	NaN
## gdp_per_capita1.161	-1.366	NaN	NaN	NaN
## gdp_per_capita1.185	0.519	NaN	NaN	NaN
## gdp_per_capita1.189	-0.451	NaN	NaN	NaN
## gdp_per_capita1.198	-0.580	NaN	NaN	NaN
## gdp_per_capita1.218	0.374	NaN	NaN	NaN
## gdp_per_capita1.223	-0.787	NaN	NaN	NaN
## gdp_per_capita1.234	0.372	NaN	NaN	NaN
## gdp_per_capita1.253	0.572	NaN	NaN	NaN
## gdp_per_capita1.261	-0.230	NaN	NaN	NaN
## gdp_per_capita1.282	-0.117	NaN	NaN	NaN
## gdp_per_capita1.285	-0.261	NaN	NaN	NaN
## gdp_per_capita1.286	-0.756	NaN	NaN	NaN
## gdp_per_capita1.289	-0.853	NaN	NaN	NaN
## gdp_per_capita1.291	0.004	NaN	NaN	NaN
## gdp_per_capita1.292	-0.107	NaN	NaN	NaN
## gdp_per_capita1.315	-0.178	NaN	NaN	NaN
## gdp_per_capita1.321	-0.469	NaN	NaN	NaN
## gdp_per_capita1.325	0.018	NaN	NaN	NaN
## gdp_per_capita1.341	-0.322	NaN	NaN	NaN
## gdp_per_capita1.343	0.447	NaN	NaN	NaN
## gdp_per_capita1.347	-0.270	NaN	NaN	NaN
## gdp_per_capita1.353	0.529	NaN	NaN	NaN
## gdp_per_capita1.356	-0.459	NaN	NaN	NaN
## gdp_per_capita1.361	0.088	NaN	NaN	NaN
## gdp_per_capita1.375	1.133	NaN	NaN	NaN
## gdp_per_capita1.384	0.323	NaN	NaN	NaN
## gdp_per_capita1.395	-0.116	NaN	NaN	NaN
## gdp_per_capita1.402	-0.242	NaN	NaN	NaN
## gdp_per_capita1.406	1.234	NaN	NaN	NaN
## gdp_per_capita1.417	-0.160	NaN	NaN	NaN
## gdp_per_capita1.431	0.362	NaN	NaN	NaN
## gdp_per_capita1.434	0.342	NaN	NaN	NaN
## gdp_per_capita1.442	0.634	NaN	NaN	NaN
## gdp_per_capita1.444	1.389	NaN	NaN	NaN
## gdp_per_capita1.464	0.811	NaN	NaN	NaN
## gdp_per_capita1.479	1.236	NaN	NaN	NaN
## gdp_per_capita1.481	1.424	NaN	NaN	NaN
## gdp_per_capita1.482	1.442	NaN	NaN	NaN
## gdp_per_capita1.484	1.204	NaN	NaN	NaN
## gdp_per_capita1.487	0.926	NaN	NaN	NaN

## gdp_per_capita1.488	0.871	NaN	NaN	NaN
## gdp_per_capita1.494	1.204	NaN	NaN	NaN
## gdp_per_capita1.504	1.297	NaN	NaN	NaN
## gdp_per_capita1.531	0.264	NaN	NaN	NaN
## gdp_per_capita1.536	0.897	NaN	NaN	NaN
## gdp_per_capita1.546	0.913	NaN	NaN	NaN
## gdp_per_capita1.552	-0.608	NaN	NaN	NaN
## gdp_per_capita1.565	1.414	NaN	NaN	NaN
## gdp_per_capita1.616	1.457	NaN	NaN	NaN
## gdp_per_capita1.626	0.568	NaN	NaN	NaN
## gdp_per_capita1.633	0.025	NaN	NaN	NaN
## gdp_per_capita1.692	0.492	NaN	NaN	NaN
## gdp_per_capita1.742	0.783	NaN	NaN	NaN
## gdp_per_capita1.871	0.295	NaN	NaN	NaN
## social_support0.000	NA	NA	NA	NA
## social_support0.396	NA	NA	NA	NA
## social_support0.432	NA	NA	NA	NA
## social_support0.435	NA	NA	NA	NA
## social_support0.513	NA	NA	NA	NA
## social_support0.571	-0.993	NaN	NaN	NaN
## social_support0.582	NA	NA	NA	NA
## social_support0.601	NA	NA	NA	NA
## social_support0.630	-1.375	NaN	NaN	NaN
## social_support0.640	0.132	NaN	NaN	NaN
## social_support0.673	NA	NA	NA	NA
## social_support0.712	-0.877	NaN	NaN	NaN
## social_support0.721	NA	NA	NA	NA
## social_support0.735	NA	NA	NA	NA
## social_support0.754	NA	NA	NA	NA
## social_support0.775	NA	NA	NA	NA
## social_support0.791	NA	NA	NA	NA
## social_support0.804	-0.618	NaN	NaN	NaN
## social_support0.832	NA	NA	NA	NA
## social_support0.865	NA	NA	NA	NA
## social_support0.868	NA	NA	NA	NA
## social_support0.871	NA	NA	NA	NA
## social_support0.872	NA	NA	NA	NA
## social_support0.874	NA	NA	NA	NA
## social_support0.905	NA	NA	NA	NA
## social_support0.913	NA	NA	NA	NA
## social_support0.935	NA	NA	NA	NA
## social_support0.946	NA	NA	NA	NA
## social_support0.954	NA	NA	NA	NA
## social_support0.968	NA	NA	NA	NA
## social_support0.979	NA	NA	NA	NA
## social_support0.984	NA	NA	NA	NA
## social_support0.993	NA	NA	NA	NA
## social_support0.997	NA	NA	NA	NA
## social_support1.003	NA	NA	NA	NA
## social_support1.006	NA	NA	NA	NA
## social_support1.007	NA	NA	NA	NA
## social_support1.042	NA	NA	NA	NA
## social_support1.043	NA	NA	NA	NA
## social_support1.068	NA	NA	NA	NA

## social_support1.069	NA	NA	NA	NA
## social_support1.081	NA	NA	NA	NA
## social_support1.083	NA	NA	NA	NA
## social_support1.098	NA	NA	NA	NA
## social_support1.104	-1.635	NaN	NaN	NaN
## social_support1.123	NA	NA	NA	NA
## social_support1.128	NA	NA	NA	NA
## social_support1.130	NA	NA	NA	NA
## social_support1.131	NA	NA	NA	NA
## social_support1.138	NA	NA	NA	NA
## social_support1.144	NA	NA	NA	NA
## social_support1.146	NA	NA	NA	NA
## social_support1.150	-0.629	NaN	NaN	NaN
## social_support1.152	NA	NA	NA	NA
## social_support1.156	-1.035	NaN	NaN	NaN
## social_support1.161	NA	NA	NA	NA
## social_support1.179	NA	NA	NA	NA
## social_support1.180	NA	NA	NA	NA
## social_support1.182	NA	NA	NA	NA
## social_support1.186	NA	NA	NA	NA
## social_support1.190	NA	NA	NA	NA
## social_support1.208	NA	NA	NA	NA
## social_support1.210	NA	NA	NA	NA
## social_support1.211	NA	NA	NA	NA
## social_support1.214	NA	NA	NA	NA
## social_support1.216	NA	NA	NA	NA
## social_support1.219	NA	NA	NA	NA
## social_support1.222	NA	NA	NA	NA
## social_support1.228	NA	NA	NA	NA
## social_support1.229	NA	NA	NA	NA
## social_support1.238	NA	NA	NA	NA
## social_support1.239	NA	NA	NA	NA
## social_support1.252	NA	NA	NA	NA
## social_support1.254	NA	NA	NA	NA
## social_support1.256	NA	NA	NA	NA
## social_support1.258	NA	NA	NA	NA
## social_support1.260	NA	NA	NA	NA
## social_support1.263	NA	NA	NA	NA
## social_support1.266	NA	NA	NA	NA
## social_support1.271	NA	NA	NA	NA
## social_support1.272	NA	NA	NA	NA
## social_support1.274	NA	NA	NA	NA
## social_support1.277	NA	NA	NA	NA
## social_support1.281	NA	NA	NA	NA
## social_support1.284	NA	NA	NA	NA
## social_support1.285	NA	NA	NA	NA
## social_support1.286	NA	NA	NA	NA
## social_support1.287	NA	NA	NA	NA
## social_support1.323	-0.864	NaN	NaN	NaN
## social_support1.338	NA	NA	NA	NA
## social_support1.340	NA	NA	NA	NA
## social_support1.343	NA	NA	NA	NA
## social_support1.354	NA	NA	NA	NA
## social_support1.358	NA	NA	NA	NA

## social_support1.367	-0.707	NaN	NaN	NaN
## social_support1.368	NA	NA	NA	NA
## social_support1.373	NA	NA	NA	NA
## social_support1.376	NA	NA	NA	NA
## social_support1.380	NA	NA	NA	NA
## social_support1.384	NA	NA	NA	NA
## social_support1.385	NA	NA	NA	NA
## social_support1.388	NA	NA	NA	NA
## social_support1.394	NA	NA	NA	NA
## social_support1.395	NA	NA	NA	NA
## social_support1.402	NA	NA	NA	NA
## social_support1.405	NA	NA	NA	NA
## social_support1.412	NA	NA	NA	NA
## social_support1.416	NA	NA	NA	NA
## social_support1.420	NA	NA	NA	NA
## social_support1.426	1.174	NaN	NaN	NaN
## social_support1.429	NA	NA	NA	NA
## social_support1.431	NA	NA	NA	NA
## social_support1.434	NA	NA	NA	NA
## social_support1.436	NA	NA	NA	NA
## social_support1.440	NA	NA	NA	NA
## social_support1.445	NA	NA	NA	NA
## social_support1.446	NA	NA	NA	NA
## social_support1.453	NA	NA	NA	NA
## social_support1.458	NA	NA	NA	NA
## social_support1.460	NA	NA	NA	NA
## social_support1.462	NA	NA	NA	NA
## social_support1.469	NA	NA	NA	NA
## social_support1.473	NA	NA	NA	NA
## social_support1.474	NA	NA	NA	NA
## social_support1.477	NA	NA	NA	NA
## social_support1.478	NA	NA	NA	NA
## social_support1.481	NA	NA	NA	NA
## social_support1.488	NA	NA	NA	NA
## social_support1.493	NA	NA	NA	NA
## social_support1.496	NA	NA	NA	NA
## social_support1.505	NA	NA	NA	NA
## social_support1.507	NA	NA	NA	NA
## social_support1.510	NA	NA	NA	NA
## social_support1.517	NA	NA	NA	NA
## social_support1.532	NA	NA	NA	NA
## social_support1.534	NA	NA	NA	NA
## social_support1.540	NA	NA	NA	NA
## social_support1.548	NA	NA	NA	NA
## social_support1.549	NA	NA	NA	NA
## social_support1.551	NA	NA	NA	NA
## social_support1.558	NA	NA	NA	NA
## social_support1.611	NA	NA	NA	NA
## healthy_life_expectancy0.000	NA	NA	NA	NA
## healthy_life_expectancy0.006	NA	NA	NA	NA
## healthy_life_expectancy0.019	NA	NA	NA	NA
## healthy_life_expectancy0.041	NA	NA	NA	NA
## healthy_life_expectancy0.049	NA	NA	NA	NA
## healthy_life_expectancy0.050	NA	NA	NA	NA

[illegible]

[illegible]


```
## healthy_life_expectancy0.797      NA      NA      NA      NA
## healthy_life_expectancy0.799      NA      NA      NA      NA
## healthy_life_expectancy0.805      NA      NA      NA      NA
## healthy_life_expectancy0.809      NA      NA      NA      NA
## healthy_life_expectancy0.810      NA      NA      NA      NA
## healthy_life_expectancy0.811      NA      NA      NA      NA
## healthy_life_expectancy0.815      NA      NA      NA      NA
## healthy_life_expectancy0.817      NA      NA      NA      NA
## healthy_life_expectancy0.818      NA      NA      NA      NA
## healthy_life_expectancy0.819      NA      NA      NA      NA
## healthy_life_expectancy0.822      NA      NA      NA      NA
## healthy_life_expectancy0.831      NA      NA      NA      NA
## healthy_life_expectancy0.834      NA      NA      NA      NA
## healthy_life_expectancy0.835      NA      NA      NA      NA
## healthy_life_expectancy0.838      NA      NA      NA      NA
## healthy_life_expectancy0.844      NA      NA      NA      NA
## healthy_life_expectancy0.845      NA      NA      NA      NA
## healthy_life_expectancy0.853      NA      NA      NA      NA
## healthy_life_expectancy0.858      NA      NA      NA      NA
## healthy_life_expectancy0.889      NA      NA      NA      NA
## healthy_life_expectancy0.900      NA      NA      NA      NA
## healthy_life_expectancy0.913      NA      NA      NA      NA
## healthy_life_expectancy0.943      NA      NA      NA      NA
## healthy_life_expectancy0.949      NA      NA      NA      NA
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:      1, Adjusted R-squared:      NaN
## F-statistic:      NaN on 155 and 0 DF, p-value: NA
```

STEP 6: Data Viz

Loading library

Notes: This code loads the “ggplot2” library, which is widely used for creating data visualizations in R. It provides a range of functions and features for creating various types of plots and charts.

```
library(ggplot2)
```

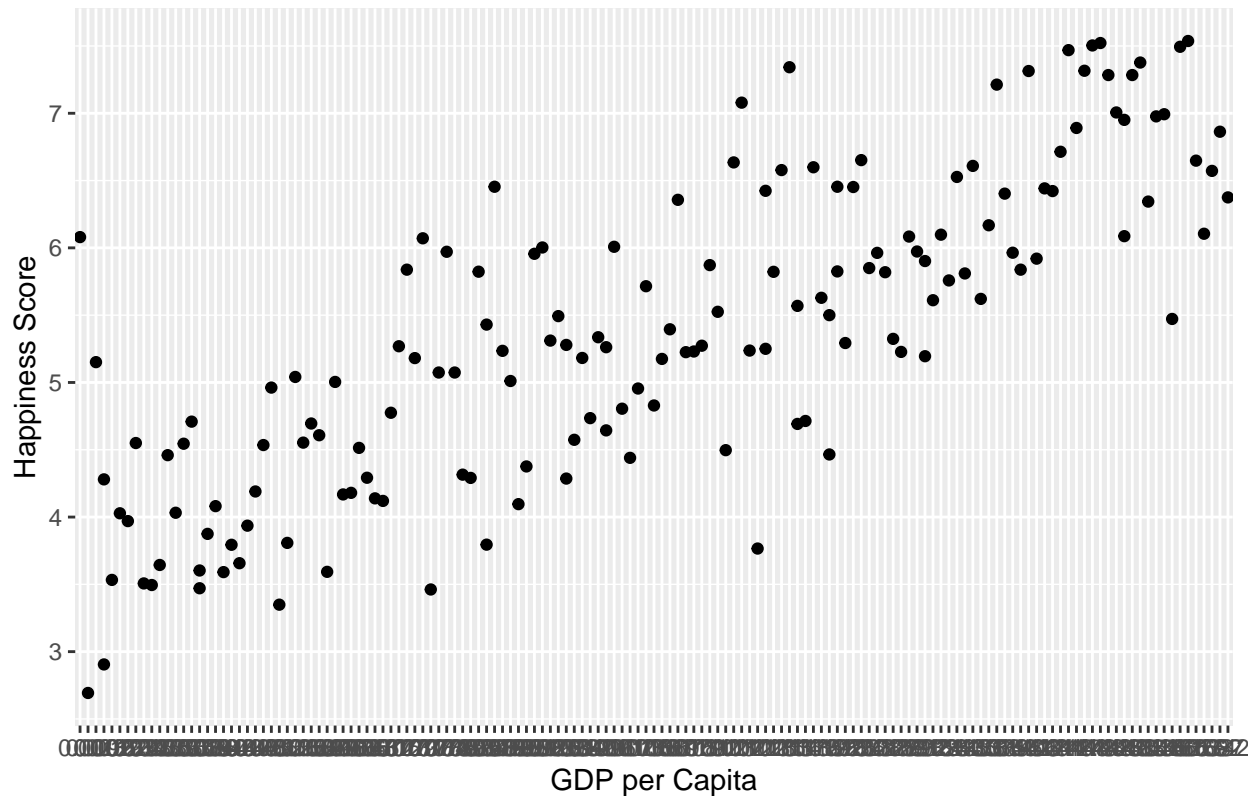
Scatter plot with a regression line

Notes: This code generates a scatter plot using “ggplot2” with GDP per capita on the x-axis and happiness score on the y-axis. Additionally, it adds a linear regression line for visualizing the relationship between the variables, accompanied by appropriate labels and a title.

```
ggplot(table_data, aes(x = gdp_per_capita, y = score)) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(title = "Scatter Plot of GDP per Capita vs. Happiness Score", x = "GDP per Capita", y = "Happiness Score")

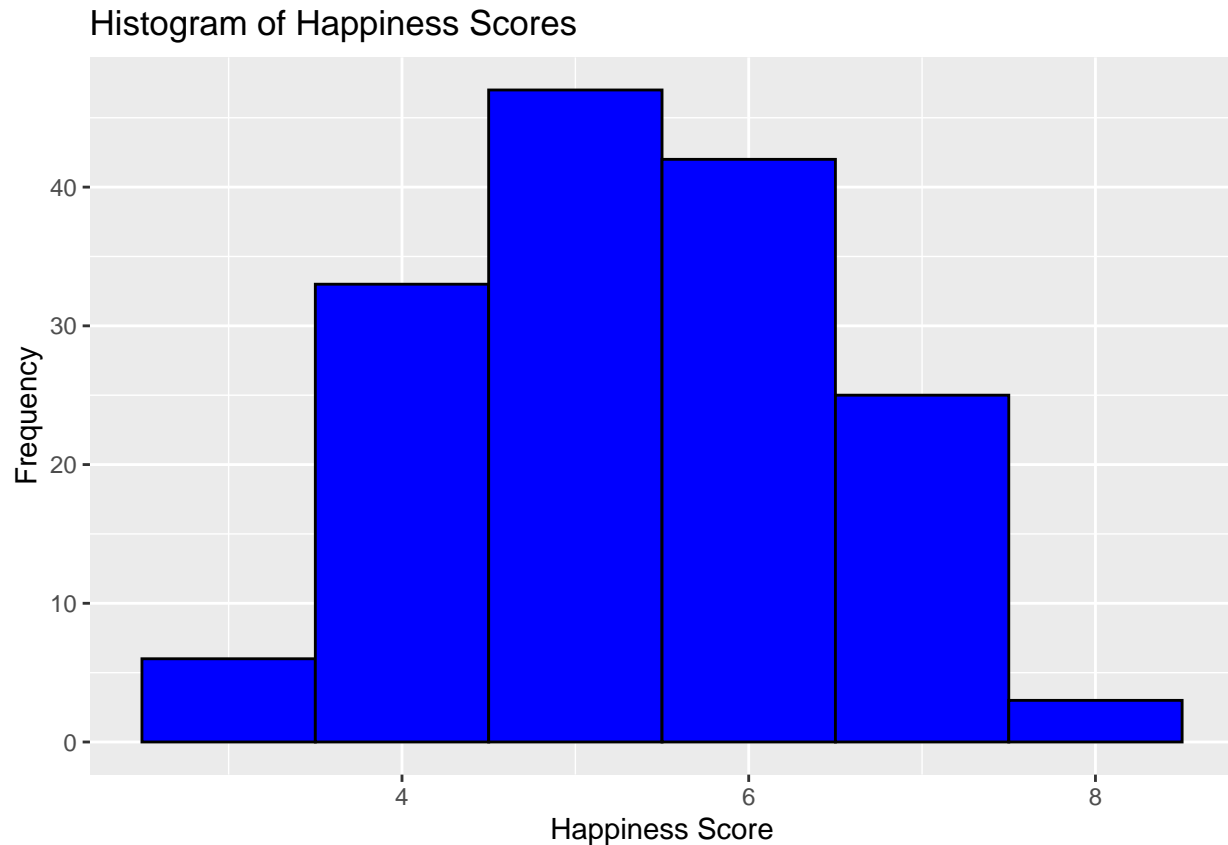
## `geom_smooth()` using formula = 'y ~ x'
```

Scatter Plot of GDP per Capita vs. Happiness Score



Histogram of the distribution of happiness scores Notes: This code creates a histogram using “ggplot2” to display the distribution of happiness scores in the “table_data” dataset. It uses a binwidth of 1 and customizes the plot with title and axis labels.

```
ggplot(table_data, aes(x = score)) +  
  geom_histogram(binwidth = 1, fill = "blue", color = "black") +  
  labs(title = "Histogram of Happiness Scores", x = "Happiness Score", y = "Frequency")
```



Heatmap to visualize the correlation matrix Notes: This code segment creates a heatmap to visualize the correlation matrix. It uses the “reshape2” package to convert the matrix into a tidy format suitable for plotting. The heatmap is generated using “ggplot2,” presenting the correlations between variables in the “table_data” dataset, with colors ranging from red (negative correlation) to green (positive correlation).

```
#install.packages("reshape2")
#library(reshape2)

#correlation_matrix <- cor(table_data %>%
#                               select(score, gdp_per_capita, social_support, healthy_life_expectancy, fre

#correlation_data <- melt(correlation_matrix)

#ggplot(correlation_data, aes(Var1, Var2, fill = value)) +
#  geom_tile() +
#  scale_fill_gradient(low = "red", high = "green") +
#  labs(title = "Correlation Heatmap", x = "Variables", y = "Variables")
```