HW 5

Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see https://quarto.org.

Running Code

When you click the **Render** button a document will be generated that includes both content and the output of embedded code. You can embed code like this:

```
1 + 1
```

```
[1] 2
```

You can add options to executable code like this

```
[1] 4
```

The `echo: false` option disables the printing of code (only output is displayed).

Author: Benjamin Cook

Class: POLI502

Date: [9/25 - 9/26]

Check dimensions of the dataset (number of rows and columns)

dim(world.data)

View the first few rows

head(world.data)

View the last few rows

tail(world.data)

Get the names of all the variables (columns)

names(world.data)

Alternatively, get the column names

colnames(world.data)

Get a summary of all variables in the dataset

summary(world.data)

Get the structure of the dataset (data types of each variable)

str(world.data)

---

2. Summarizing Categorical Variables

table(world.data$democ_regime)

table(world.data$typerel)

ft.typerel <- data.frame(table(world.data$typerel)) ft.typere

ft.colony <- data.frame(table(world.data $ colony)) ft.colony

sum(ft.colony $ Freq)

ft.colony $ Freq / sum(ft.colony $ Freq)

prop.table(ft.colony $ Freq)

prop.table(ft.colony $ Freq) * 100

round(prop.table(ft.colony $ Freq) * 100, digits = 2) ft.colony

ft.colony $ Percent <- round(prop.table(ft.colony $ Freq) * 100, digits = 2) ft.colony

colnames(ft.colony)[colnames(ft.colony) == “Var1”] <- “Colonizer” ft.colony

3. Create and save a bar chart for the typerel variable

typerel_freq <- table(world.data$typerel) typerel_freq

most_popular_religion <- names(which.max(typerel_freq))

Display the most popular religion

most_popular_religion # “Roman Catholic”

#Muslim Count

muslim_count <- typerel_freq[“Muslim”] total_countries <- sum(typerel_freq) muslim_percentage <- (muslim_count / total_countries) * 100 # 26.2

Dem Regime

democratic_count <- democ_regime_freq[“Yes”]

total_countries <- sum(democ_regime_freq, na.rm = TRUE)

democratic_percentage <- (democratic_count / total_countries) * 100

60.3

Chart

ggplot(world.data, aes(x = typerel)) + geom_bar() + xlab("Predominant Religion") + ylab("Number of Countries") + ggtitle("Distribution of Predominant Religions in Countries")

g <- ggplot(world.data, aes(x = democ_regime)) + geom_bar() + xlab("Democratic Regime") + ylab("Number of Countries") + ggtitle("Distribution of Democratic Regimes in Countries")

Saving PDF

g <- ggplot(world.data, aes(x = typerel)) + geom_bar() + xlab("Predominant Religion") + ylab("Number of Countries") + ggtitle("Distribution of Predominant Religions in Countries") ggsave(filename = "typerel_bar_chart.pdf", plot = g, width = 10, height = 8)

Cent Tend

mean_gini04 <- mean(world.data$gini04, na.rm = TRUE) median_gini04 <- median(world.data$gini04, na.rm = TRUE) sd_gini04 <- sd(world.data$gini04, na.rm = TRUE) var_gini04 <- var(world.data$gini04, na.rm = TRUE) range_gini04 <- range(world.data$gini04, na.rm = TRUE) summary_gini04 <- list(Mean = mean_gini04, Median = median_gini04, Standard_Deviation = sd_gini04, Variance = var_gini04, Range = range_gini04)

Display the summary

summary_gini04

#Interpreting Central Tendency:

#If the mean and median of gini08 are higher than those of gini04, it suggests that economic inequality is getting worse because, on average, countries are more unequal in 2008 than they were in 2004. #If the mean and median are lower in 2008, it suggests that economic inequality is improving

#Histogram

hist_gini04 <- ggplot(world.data, aes(x = gini04)) + geom_histogram(binwidth = 0.05, fill = “blue”, color = “black”) + xlab(“Gini Coefficient (2004)”) + ylab(“Number of Countries”) + ggtitle(“Distribution of Gini Coefficient in 2004”) hist_gini04 ggsave(filename = “gini04_histogram.pdf”, plot = hist_gini04, width = 10, height = 8)

hist_gini08 <- ggplot(world.data, aes(x = gini08)) + geom_histogram(binwidth = 0.05, fill = “green”, color = “black”, na.rm = TRUE) + xlab(“Gini Coefficient (2008)”) + ylab(“Number of Countries”) + ggtitle(“Distribution of Gini Coefficient in 2008”)

hist_gini08

ggsave(filename = “gini08_histogram.pdf”, plot = hist_gini08, width = 10, height = 8)

#Income Dist. is getting worse.

Facet Wrap

hist_gini04 <- ggplot(world.data, aes(x = gini04)) + geom_histogram(binwidth = 0.05, fill = “blue”, color = “black”, na.rm = TRUE) + xlab(“Gini Coefficient (2004)”) + ylab(“Number of Countries”) + ggtitle(“Distribution of Gini Coefficient in 2004”)

hist_gini04_by_region <- hist_gini04 + facet_wrap(~ region) hist_gini04_by_region

#Women

hist_women09_by_region <- ggplot(world.data, aes(x = women09)) + geom_histogram(binwidth = 5, fill = “purple”, color = “black”, na.rm = TRUE) + xlab(“Percentage of Women in Parliament (2009)”) + ylab(“Number of Countries”) + ggtitle(“Distribution of Women in Parliament in 2009 by Region”) + facet_wrap(~ region) hist_women09_by_region

gen4 SD

sd_gini04_by_region <- by(world.data$gini04, world.data$region, function(x) sd(x, na.rm = TRUE)) sd_gini04_by_region

#Smallest Dis world.data$region: Scandinavia #[1] 0.9831921