PLS 397: Midterm

Zane Shango

2/25/2020

Guidelines

Midterm is open book and open internet but you must work alone. You must submit your .Rmd and a .PDF version of it to the midterm d2l folder by **3/3 12pm**. The data file you will be using for the midterm is: `midtermData.rda` .

```
rm(list=ls())

# loading in the .rda below will add an object to your workspace called data
load('midtermData.rda')

# data is a tibble dataframe
str(data) # structure of data
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    5574 obs. of  10 variables:
##  $ country   : chr  "Afghanistan" "Albania" "Albania" "Albania" ...
##  $ region    : Factor w/ 7 levels "Middle East & North Africa",..: 2 3 3 3 3 3 3 3 3
3 ...
##  $ year      : int  2014 1981 1980 1988 1989 1990 1987 1993 1994 1982 ...
##  $ gdpCap    : num  584 2065 1992 1936 2070 ...
##  $ lifeExpect: num  63 70.4 70.2 71.8 71.9 ...
##  $ pop       : num  33370794 2726056 2671997 3142336 3227943 ...
##  $ logGdpCap : num  6.37 7.63 7.6 7.57 7.64 ...
##  $ logPop    : num  17.3 14.8 14.8 15 15 ...
##  $ polity2   : int  -1 -9 -9 -9 -9 1 -9 5 5 -9 ...
##  $ co2       : num  0.294 2.693 1.935 2.332 2.783 ...
##  - attr(*, "na.action")= 'omit' Named int  1 2 3 4 5 6 7 9 10 11 ...
##   ..- attr(*, "names")= chr  "1" "2" "3" "4" ...
```

```
# we have both information from wBank and polity
# world bank variables:
  # gdpCap (GDP per capita (constant 2000 USD))
  # lifeExpect (life expectancy (years))
  # pop (population)
  # logGdpCap (logged GDP per capita (constant 2000 USD))
  # logPop (logged population)
  # co2 (carbon dioxide emissions in metric tons per capita)
# polity variables:
  # polity2: -10 to 10, -10 indicates full autocracy, 10 indicates full democracy
```

Additionally, you will want to load the tidyverse library.

```
# loading the tidyverse library will also call the
# ggplot2, dplyr, and magrittr (%>%) libraries to your workspace
library(tidyverse)
```

Q1a: Plotting

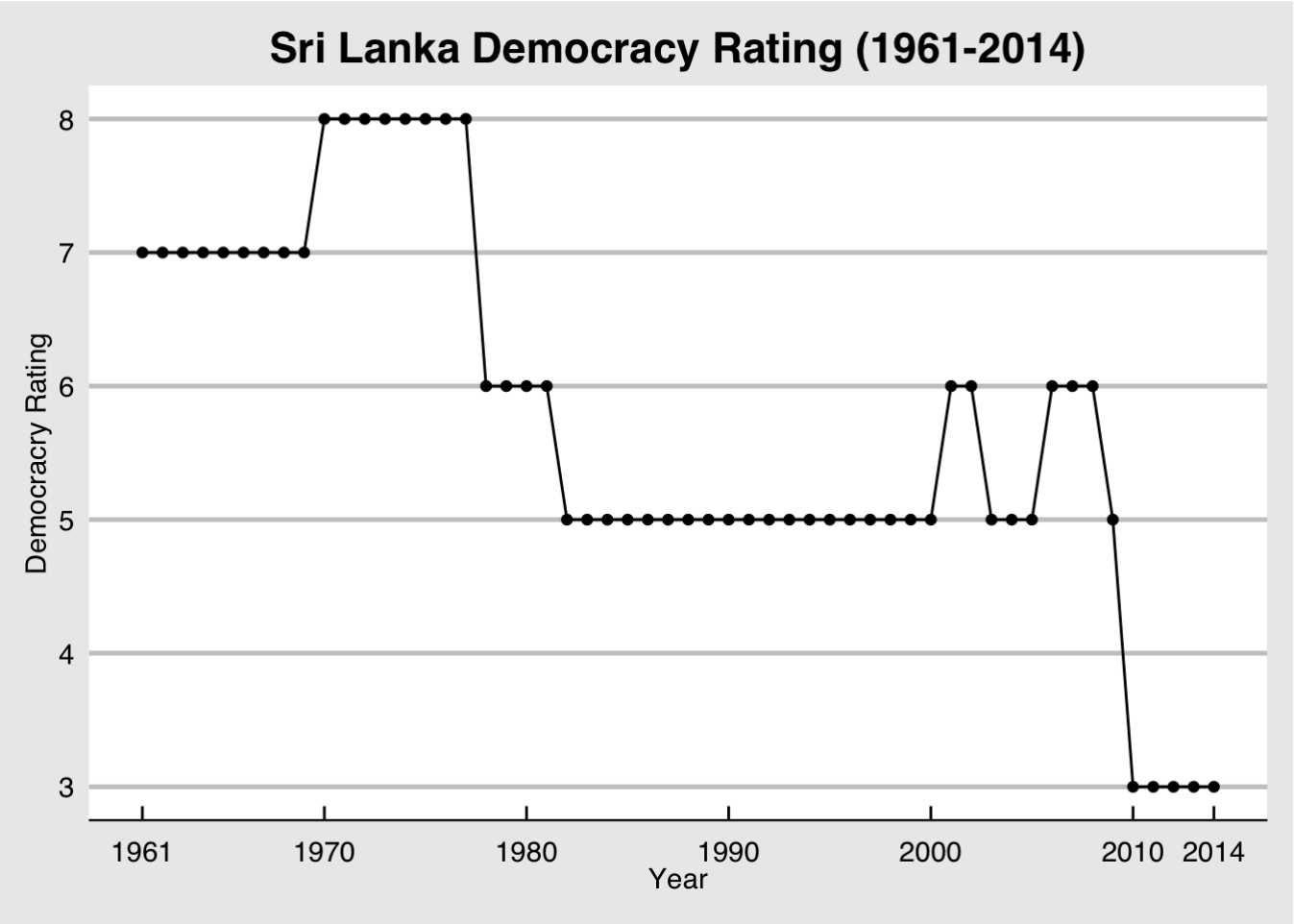
```
library(ggthemes)

sri_lanka <- data %>%
  filter((country == "Sri Lanka") & (year >= 1960 & year <= 2016))

# min and max of years (range) that Sri Lanka has ratings for
min(sri_lanka$year)
```

```
## [1] 1961
```

```
max(sri_lanka$year)
```

```
## [1] 2014
```

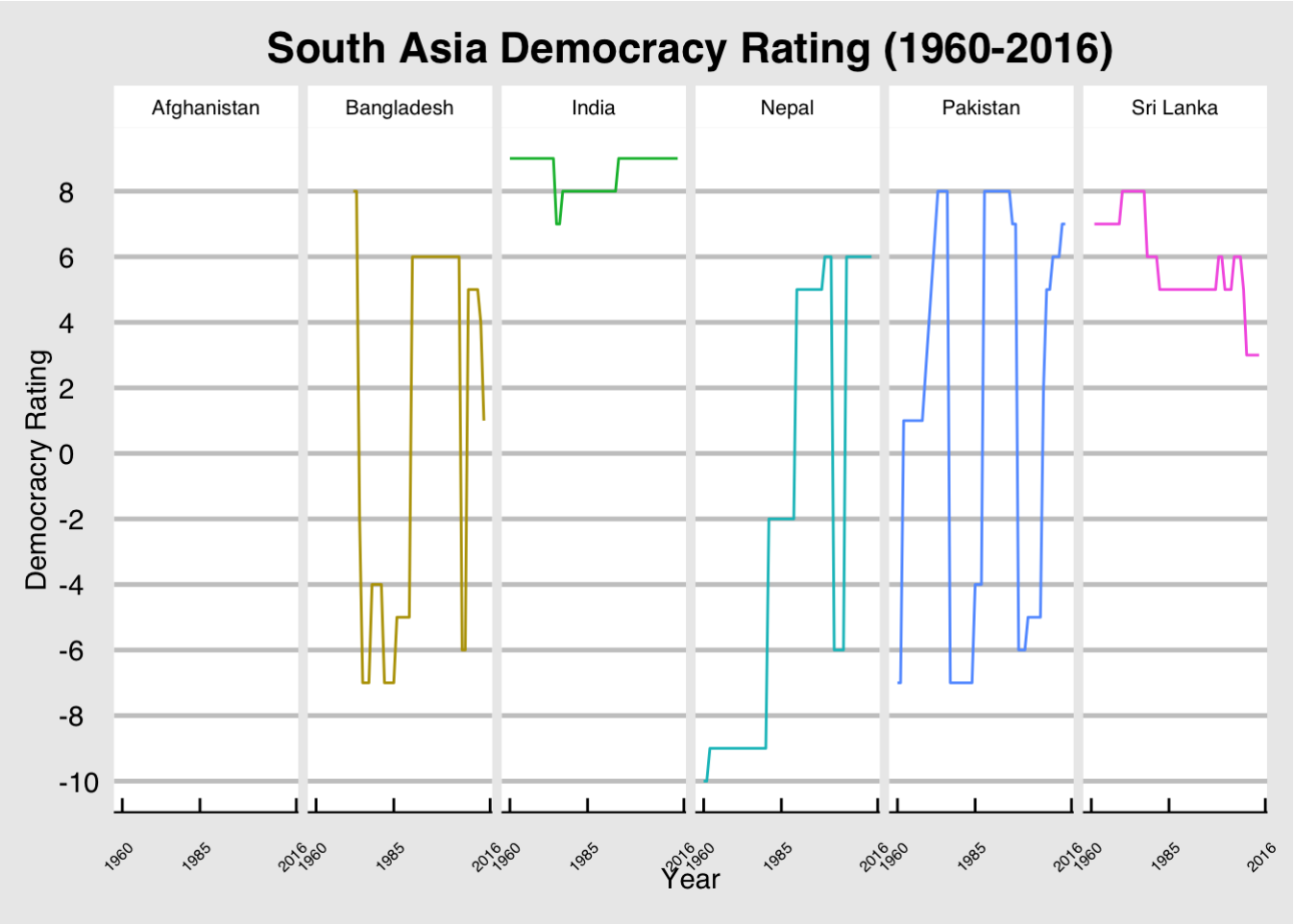
```
sri_plot <- ggplot(data = sri_lanka, aes(x = year, y = polity2)) +
  geom_point() +
  geom_line() +
  scale_x_continuous(breaks = c(1961,1970,1980,1990,2000,2010,2014)) +
  scale_y_continuous(breaks = c(1,2,3,4,5,6,7,8,9,10)) +
  labs(
    title = "Sri Lanka Democracy Rating (1961-2014)",
    x = "Year",
    y = "Democracry Rating") +
  theme_economist_white() +     # use white colored economist theme
  theme(plot.title = element_text(hjust = .5)) # override that theme for the title posit
ion to be centered

sri_plot
```



Q1b: Plotting

```
s_asia <- data %>%
  filter(region == "South Asia") %>%
  group_by(country)


s_asia_plot <- ggplot(data = s_asia, aes(x = year, y = polity2, color = country)) +
  geom_line() +
  scale_x_continuous(breaks = c(1960,1985,2016)) +
  scale_y_continuous(breaks = c(-10,-8,-6,-4,-2,0,2,4,6,8,10)) +
  labs(
    title = "South Asia Democracy Rating (1960-2016)",
    x = "Year",
    y = "Democracry Rating",
    color = "Country") +
  theme_economist_white() +
  theme(plot.title = element_text(hjust = .5),
        strip.text.x = element_text(size = 8),
        axis.text.x = element_text(size = 6, angle = 45),
        legend.position = "none") +
  facet_wrap(~country, nrow = 1)


# remove legend because the line colors already clearly correspond to the south asian co
untries
s_asia_plot
```

```
## geom_path: Each group consists of only one observation. Do you need to adjust
## the group aesthetic?
```



Sort of hard to read, but you can see India and Sri Lanka have steady democracy ratings in positive values (India is by far the most democratic). Nepal is on the rise, but is still faily inconsistent, while Bangladesh follows the opposite trend where is seems to be getting worse, but not in a consistent trend. Pakistan is all over the place and it appears that depending on the leader the country will be either very democratic or borderline autocratic, which is strange. Afghanistan only has one observed year, but is rated below 0.

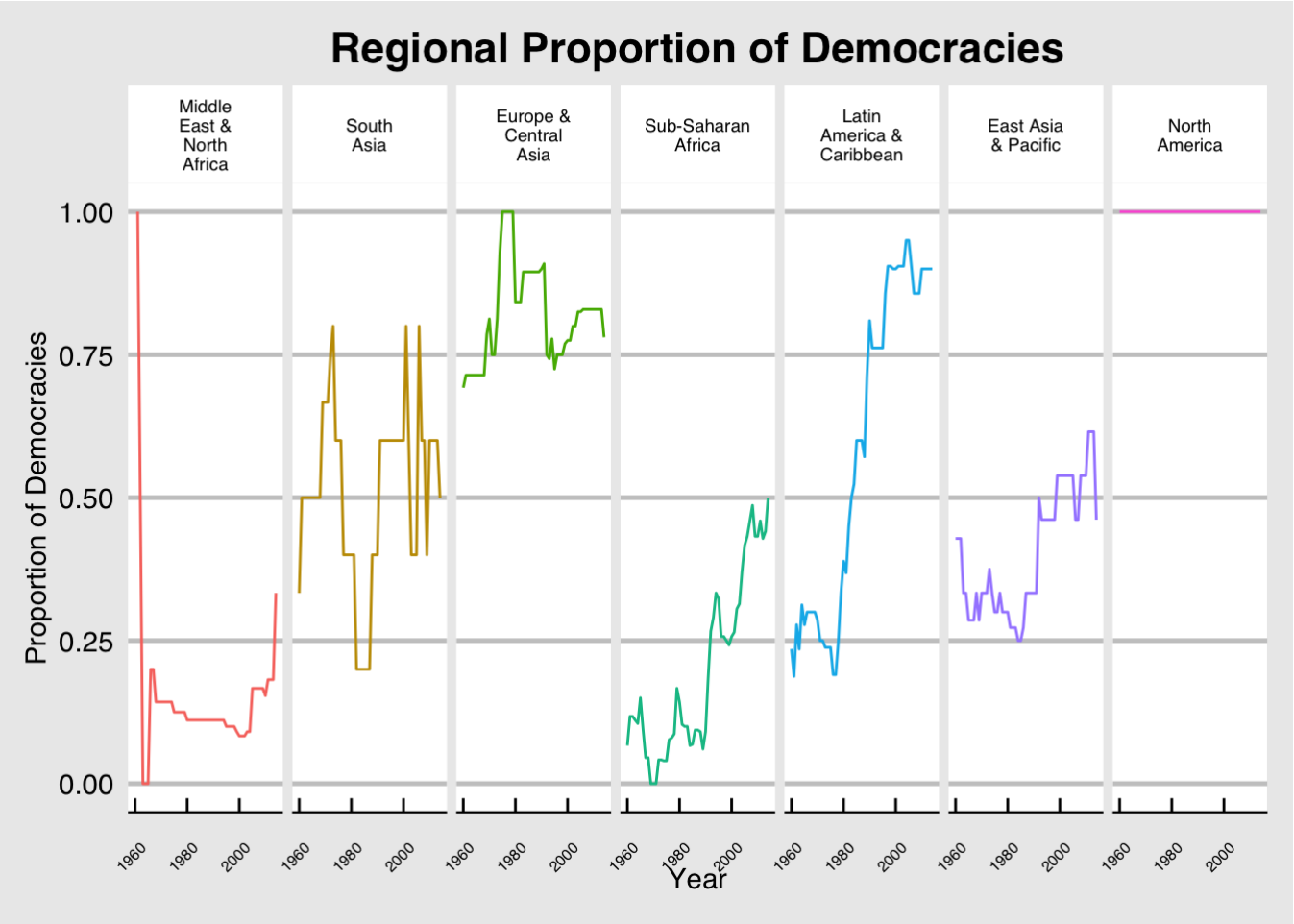
Q2a: Summarizing TSCS Data

```
# new dataframe with column giving 1 if that polity2 value is greater than or equal to 6
and 0 if not
democ_g6 <- data %>%
  mutate(democ_greater6 = ifelse(polity2 >= 6, 1, 0))

# group by region and year and get mean value of countries in that region, that year wit
h a democ rating of 6 or higher
democs_df <- democ_g6 %>%
  group_by(region,year) %>%
  summarize(mean_democs = mean(democ_greater6))

democs_df
```

```
## # A tibble: 384 x 3
## # Groups:   region [7]
##    region                       year mean_democs
##    <fct>                       <int>       <dbl>
##  1 Middle East & North Africa  1961       1
##  2 Middle East & North Africa  1962       0.5
##  3 Middle East & North Africa  1963       0
##  4 Middle East & North Africa  1964       0
##  5 Middle East & North Africa  1965       0
##  6 Middle East & North Africa  1966       0.2
##  7 Middle East & North Africa  1967       0.2
##  8 Middle East & North Africa  1968       0.143
##  9 Middle East & North Africa  1969       0.143
## 10 Middle East & North Africa  1970       0.143
## # … with 374 more rows
```

Q2b: Visualizing Summarized TSCS Data

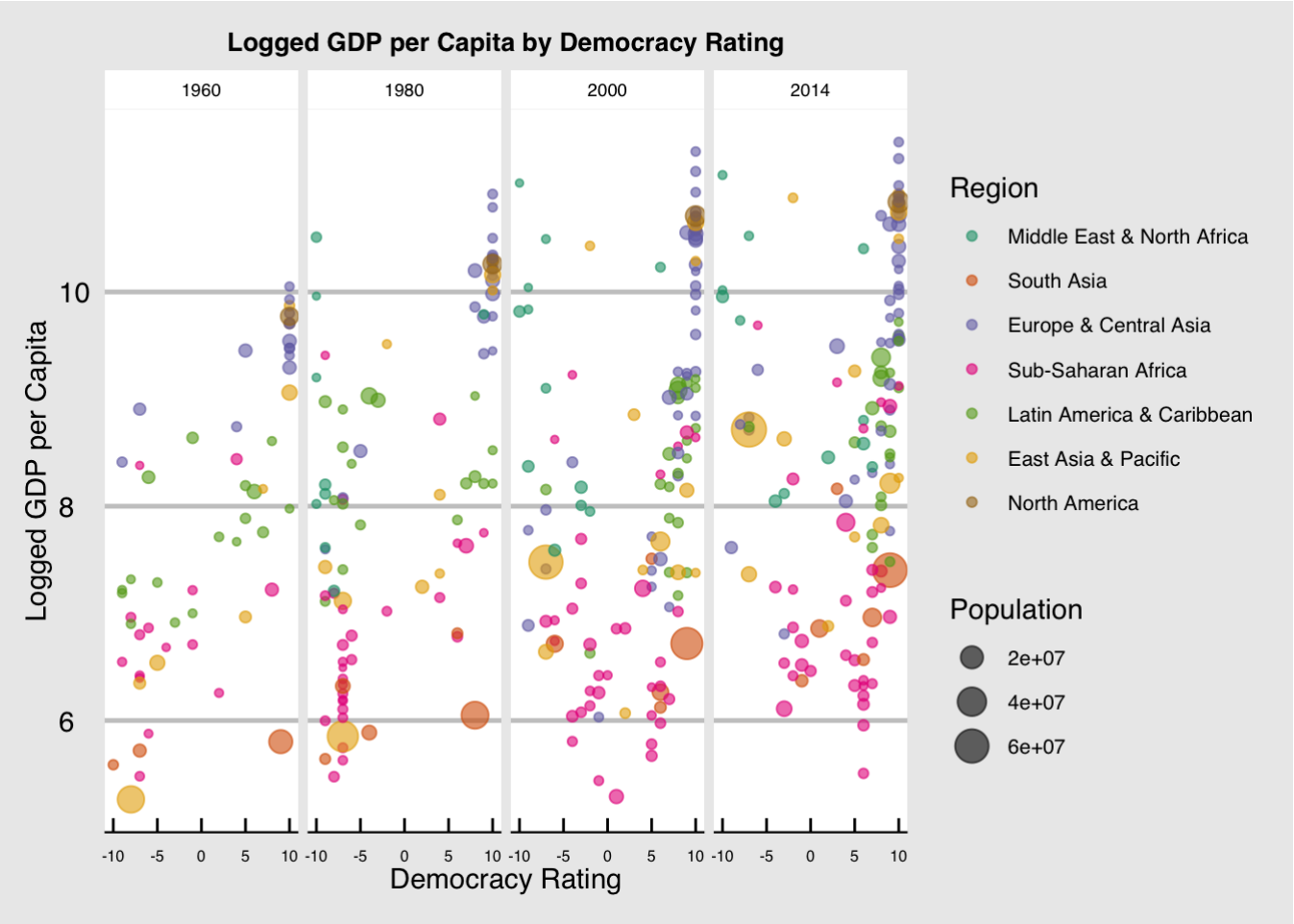
```
# visualize the number of democracies in a region changing over time

democs_region_plot <- ggplot(data = democs_df, aes(x = year, y = mean_democs, color = re
gion)) +
  geom_line() +
  labs(
    title = "Regional Proportion of Democracies",
    x = "Year",
    y = "Proportion of Democracies") +
  theme_economist_white() +
  theme(plot.title = element_text(hjust = .5),
        strip.text.x = element_text(size = 7),
        axis.text.x = element_text(size = 6, angle = 45),
        legend.position = "none") +
  facet_wrap(~region, nrow = 1,labeller = label_wrap_gen(width=10))

# labeller in facet_wrap allows me to wrap the title of the facets so it is readable
# remove legend becuase again the colors clearly are associated with regions

democs_region_plot
```



These results are mostly what I expected. North America has been all democracies since 1960, which is impressive, while Latin America and the Caribbean have been steadily rising to the point where almost all of those countries are democracies. Sub-Saharan Africa and East Asia & Pacific have both been rising as well which is good to see. South Asia has been steadily in the range of roughly half democracies, while Europe and Central Asia has been steady at above 75% of those countries have been rated democracies. The most interesting one in my opinion is the Middle East and North Africa, where in 1960 all the countries were democratic however it took a steep dive do below 25% of them being democratic for about 30 years before it's started to rise towards more democracies once again. I believe this lack of democracy presence in the Middle East would be due to the rise of many dictator-like rulers such as Saddam Hussein in Iraq, as well as the domination of Oil on the overall Middle Eastern Economy in those years.

Q3a: Income and Democracy

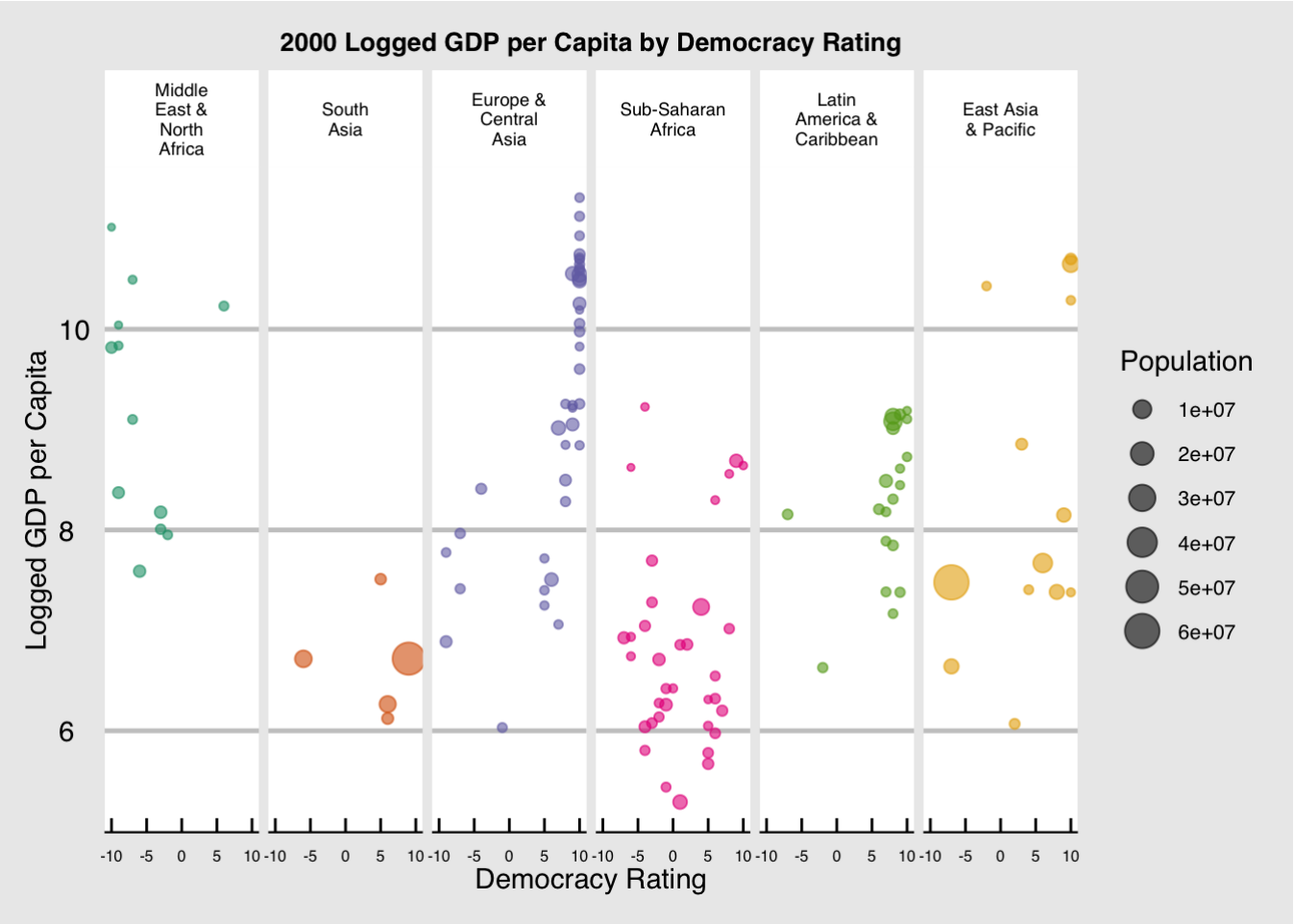
```
# create a new dataframe for plotting the relationship between GDP and democracy rating

gdp_democ <- data %>%
  filter(year %in% c(1960,1980,2000,2014)) %>%
  mutate(pop_adjust = pop/logPop)

# chose pop/logPop because it gives the most levels of population without them being sup
er close together like logPop, while pop only gives 2 levels

gdp_democ_plot <- ggplot(data = gdp_democ, aes(x = polity2, y = logGdpCap, color = regio
n, size = pop_adjust)) +
  geom_point(alpha = .6) +
  labs(
    title = "Logged GDP per Capita by Democracy Rating",
    x = "Democracy Rating",
    y = "Logged GDP per Capita",
    color = "Region",
    size = "Population") +
  theme_economist_white() +
  theme(plot.title = element_text(hjust = .5, size = 10),
        strip.text.x = element_text(size = 7),
        axis.text.x = element_text(size = 6),
        legend.position = "right",
        legend.text = element_text(size = 8)) +
  facet_wrap(~year, nrow = 1) +
  scale_color_brewer(palette = 'Dark2')

gdp_democ_plot
```



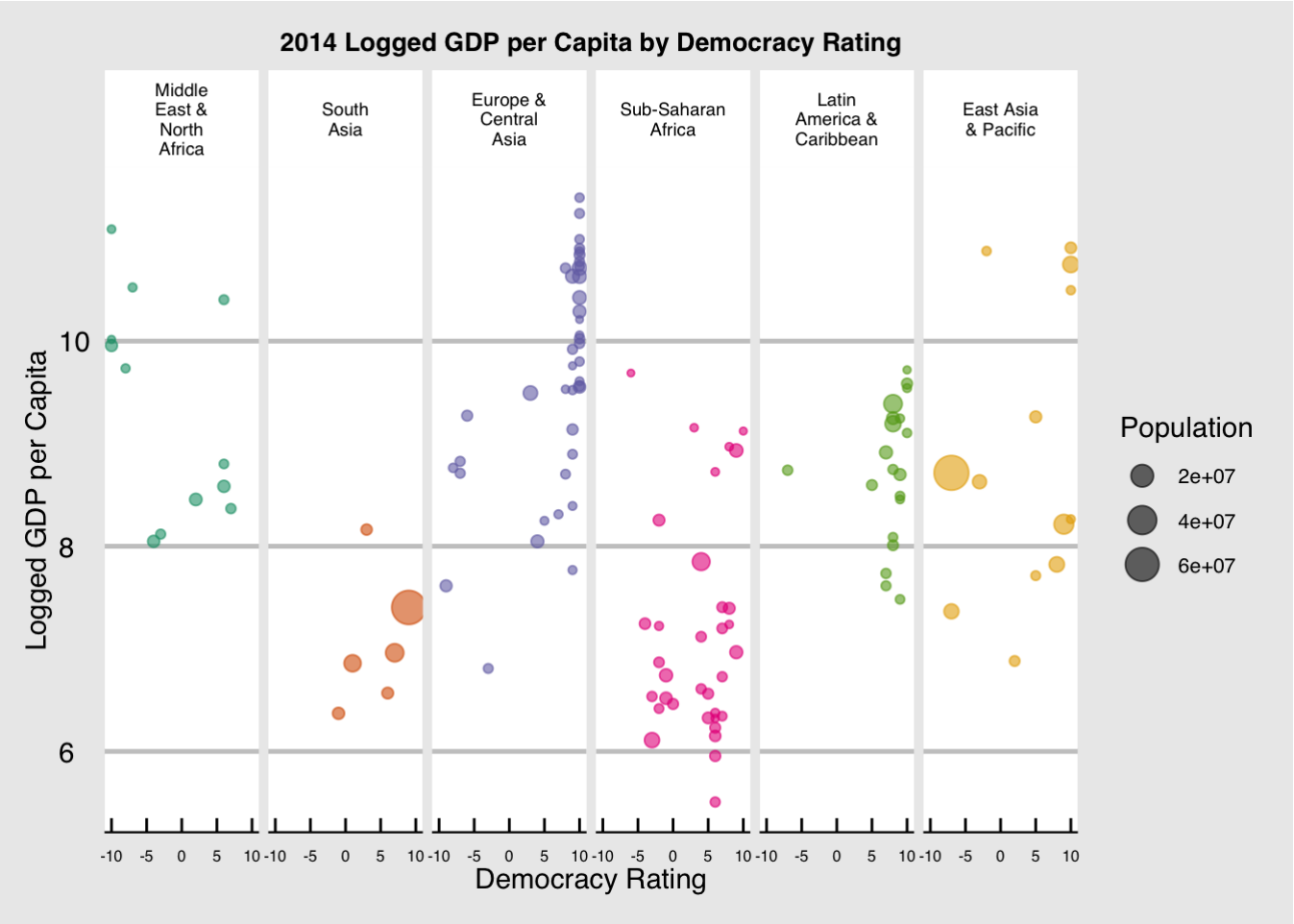
This is what I expected, showing North American and Europe & Central American to have the highest democracy ratings as which seem to be trending towards higher logged GDP per capita, which some excepts to this idea like osme Middle Eastern countries for example.

Q3b: Income and Democracy

```
# 2000 dataframe without north america
no_nAmerica00 <- data %>%
  filter((!region == "North America") & (year == 2000)) %>%
  mutate(pop_adjust = pop/logPop)

# 2000 plot
inc_dem_00 <- ggplot(data = no_nAmerica00, aes(x = polity2, y = logGdpCap, color = regio
n, size = pop/logPop)) +
  geom_point(alpha = .6) +
  labs(
    title = "2000 Logged GDP per Capita by Democracy Rating",
    x = "Democracy Rating",
    y = "Logged GDP per Capita",
    color = "Region",
    size = "Population") +
  theme_economist_white() +
  theme(plot.title = element_text(hjust = .5, size = 10),
        strip.text.x = element_text(size = 7),
        axis.text.x = element_text(size = 6),
        legend.position = "right",
        legend.text = element_text(size = 8)) +
  guides(color = FALSE) +
  facet_wrap(~region, nrow = 1, labeller = label_wrap_gen(width=10)) +
  scale_color_brewer(palette = 'Dark2')

# guides allows me to set color equal to false, so it removes the pointless region/color
legend
inc_dem_00
```




```
# 2014 dataframe without north america
no_nAmerica14 <- data %>%
  filter((!region == "North America") & (year == 2014)) %>%
  mutate(pop_adjust = pop/logPop)

# 2014 plot
inc_dem_14 <- ggplot(data = no_nAmerica14, aes(x = polity2, y = logGdpCap, color = regio
n, size = pop/logPop)) +
  geom_point(alpha = .6) +
  labs(
    title = "2014 Logged GDP per Capita by Democracy Rating",
    x = "Democracy Rating",
    y = "Logged GDP per Capita",
    color = "Region",
    size = "Population") +
  theme_economist_white() +
  theme(plot.title = element_text(hjust = .5, size = 10),
        strip.text.x = element_text(size = 7),
        axis.text.x = element_text(size = 6),
        legend.position = "right",
        legend.text = element_text(size = 8)) +
  guides(color = FALSE) +
  facet_wrap(~region, nrow = 1, labeller = label_wrap_gen(width=10)) +
  scale_color_brewer(palette = 'Dark2')

inc_dem_14
```




```
# 2014 plot
```

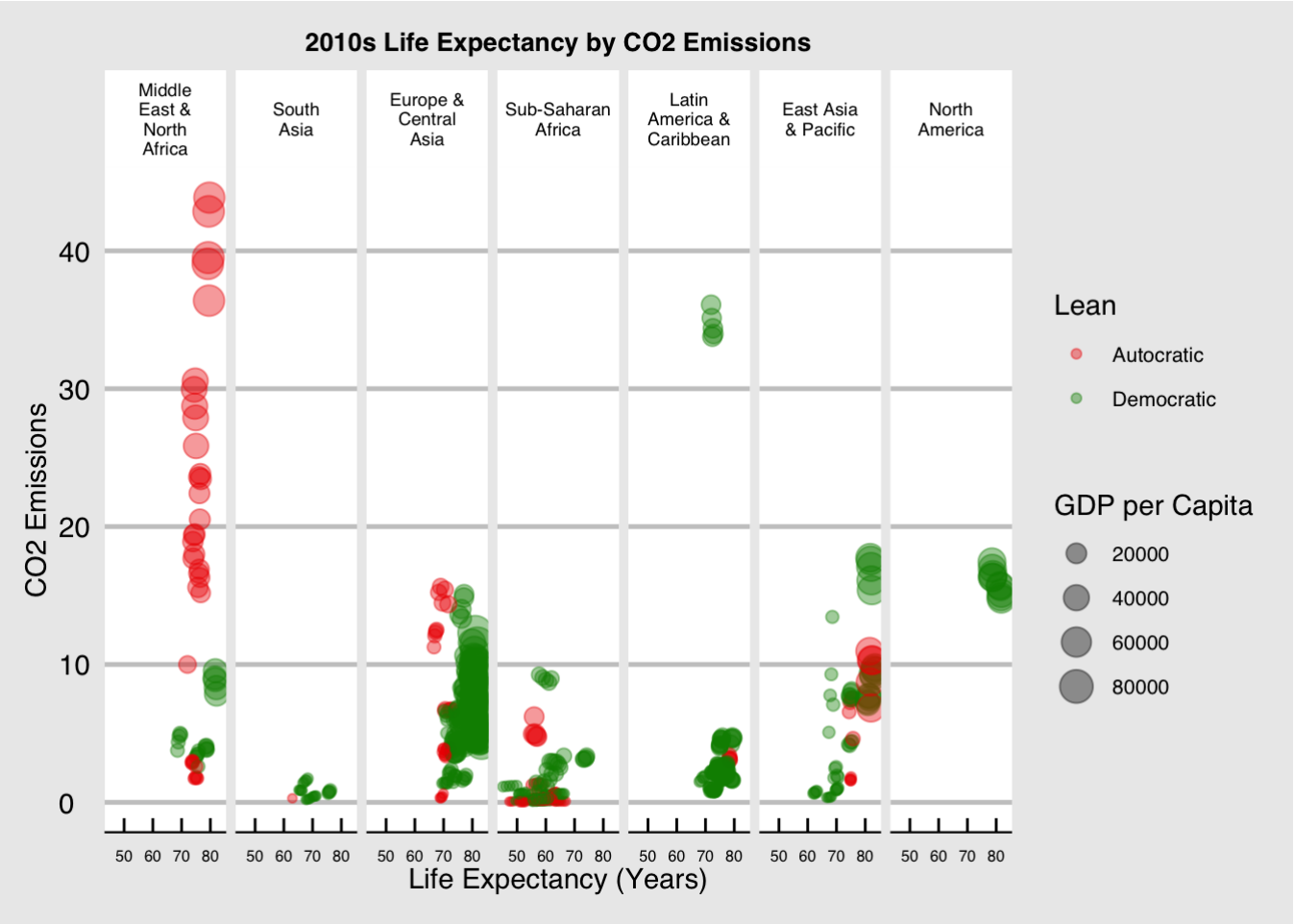
In Europe & Central Asia, a higher democracy rating seems to lead to higher logged GDP per Capita, as well as in Latin American & Caribbean and South Asia (though those aren't quite as clear). East Asia & Pacific follows this trend in a way, but it is not very clear as it is much more split. Sub-Saharan Africa doesn't seem to follow much of a trend at all, as those countries are spread out fairly evenly. However, in the Middle East & North Africa, a less democratic country can lead to a higher logged GDP per Capita, and in fact there appears to be only one country with a positive democracy rating (it ranks 3rd in logged GDP per capita). These trends above stayed just about the same in the year 2014, however Middle East & North Africa took a step in the direction of democracy, and it doesn't appear that their logGDPcap was effected much at all, which is a good sign for the future economic status of the Middle East & North Africa.

Q4

```
# I am going to focus on how c02 emissions may effect life expectancy, while attempting
 to show the affect c02 emmissions has on logGDPcap
# this is because co2 emmissions are caused mostly by industrialization that can lead to
higher GDP
# I will mutate a column to determine if the country is leaning democratic or leaning au
tocratic
# I will also only use the years between 2010-2014

q4_df <- data %>%
  filter(year %in% c(2010,2011,2012,2013,2014)) %>%
  mutate(lean = ifelse(polity2 > 0, "Democratic", "Autocratic"))

q4_plot <- ggplot(data = q4_df, aes( x = lifeExpect, y = co2, color = lean, size = gdpCa
p)) +
  geom_point(alpha = .4) +
  labs(
    title = "2010s Life Expectancy by CO2 Emissions",
    x = "Life Expectancy (Years)",
    y = "CO2 Emissions",
    color = "Lean",
    size = "GDP per Capita") +
  theme_economist_white() +
  theme(plot.title = element_text(hjust = .5, size = 10),
        strip.text.x = element_text(size = 7),
        axis.text.x = element_text(size = 6),
        legend.position = "right",
        legend.text = element_text(size = 8)) +
  facet_wrap(~region, nrow = 1, labeller = label_wrap_gen(width=10)) +
  scale_color_brewer(palette = 'Dark2') +
  scale_color_manual(values = c('Democratic' = 'green4', 'Autocratic' = 'red2'))
```

```
## Scale for 'colour' is already present. Adding another scale for 'colour',
## which will replace the existing scale.
```

```
q4_plot
```



As I expected, just about all my assumptions were found here: as GDP per Capita increased co2 emissions seemed to increase in most cases as well, the countries with more democratic leanings had higher life expectancies (except for East Asia & Pacific which had multiple autocratic countries with what appears to be some of the highest life expectancies), and as with most of the other questions the Middle East & North Africa follows opposing trends to the majority of the other regions as their mostly autocratic still (however they have by far the highest co2 emissions).