Technical Appendix

Antonio Skilton

10/12/2016

Cleaning the data

The "flash" dataset contains seven columns and 568 rows. The last two columns are manipulated in the code below so that they are more readable. Each row is a country in a given year.

Columns:

- country
- continent
- year
- life expectancy
- GDP per capita
- GDP in billions of dollars
- population in millions of people

```
read.csv("4_Flash_Proj_1_Data.csv") %>% tbl_df %>%
  mutate(country = str_replace_all(country,"[.,]",""),
         country = as.factor(country),
         gdpPercap = gdpPercap/1000,
         popMill = popThous/1000) %>%
  mutate(first_world = ifelse(continent == "Europe" |
                                continent == "Oceania" |
                                country == "United States", 1, 0),
         continent_world = ifelse(first_world == 1,
                                  "First World",
                                  as.character(continent))) %>%
  select(-popThous) -> flash
```

Table 1: First ten rows of manipulated flash data

country	continent	year	lifeExp	gdpPercap	gdpBillions	popMill	first_world	continent_world
Algeria	Africa	1972	55	4.18	61.7	14.76	0	Africa
Algeria	Africa	1977	58	4.91	84.2	17.15	0	Africa
Algeria	Africa	1982	61	5.75	115.1	20.03	0	Africa
Algeria	Africa	1987	66	5.68	132.1	23.25	0	Africa
Algeria	Africa	1992	68	5.02	132.1	26.30	0	Africa
Algeria	Africa	1997	69	4.80	139.5	29.07	0	Africa
Algeria	Africa	2002	71	5.29	165.4	31.29	0	Africa
Algeria	Africa	2007	72	6.22	207.4	33.33	0	Africa
Angola	Africa	1972	38	5.47	32.3	5.89	0	Africa
Angola	Africa	1977	39	3.01	18.5	6.16	0	Africa

Life Expectancy

According to the regression results, we see that a life expectancy is highly correlated with GDP per capita. Furthermore, we see that every continent has as higher life expectancy than Africa even when GDP per capita is held constant. Even Oceania, which is represented only by Australia, has a significantly different life expectancy than Africa.

```
flash %>%
  group_by(country,continent) %>%
  summarise_at(vars(lifeExp:popMill),mean) %>%
  lm(lifeExp ~ poly(gdpPercap,3) + continent,.) -> mod
```

Table 2:

	Dependent variable:
	lifeExp
GDP/capita	37.465^{***}
	(6.623)
$(\text{GDP/capita})^2$	-18.399^{***}
	(5.805)
$(\text{GDP/capita})^3$	14.359^{***}
	(5.053)
Americas	11.153^{***}
	(1.947)
Asia	9.051^{***}
	(1.665)
Europe	12.998^{***}
	(2.279)
Oceania	13.702^{**}
	(5.296)
Constant	56.252^{***}
	(1.075)
Observations	71
R^2	0.831
Adjusted R^2	0.812
Residual Std. Error	4.793 (df = 63)
F Statistic	44.292^{***} (df = 7; 63)
Note:	*p<0.1; **p<0.05; ***p<0.01

Here I create 'flashForPlot', a dataset where the average life expectancy of all five years is taken. 'flashForPlot' is designed to be used by ggplot2. It adds a column named 'fitted', which is the fitted values from the linear regression model above.

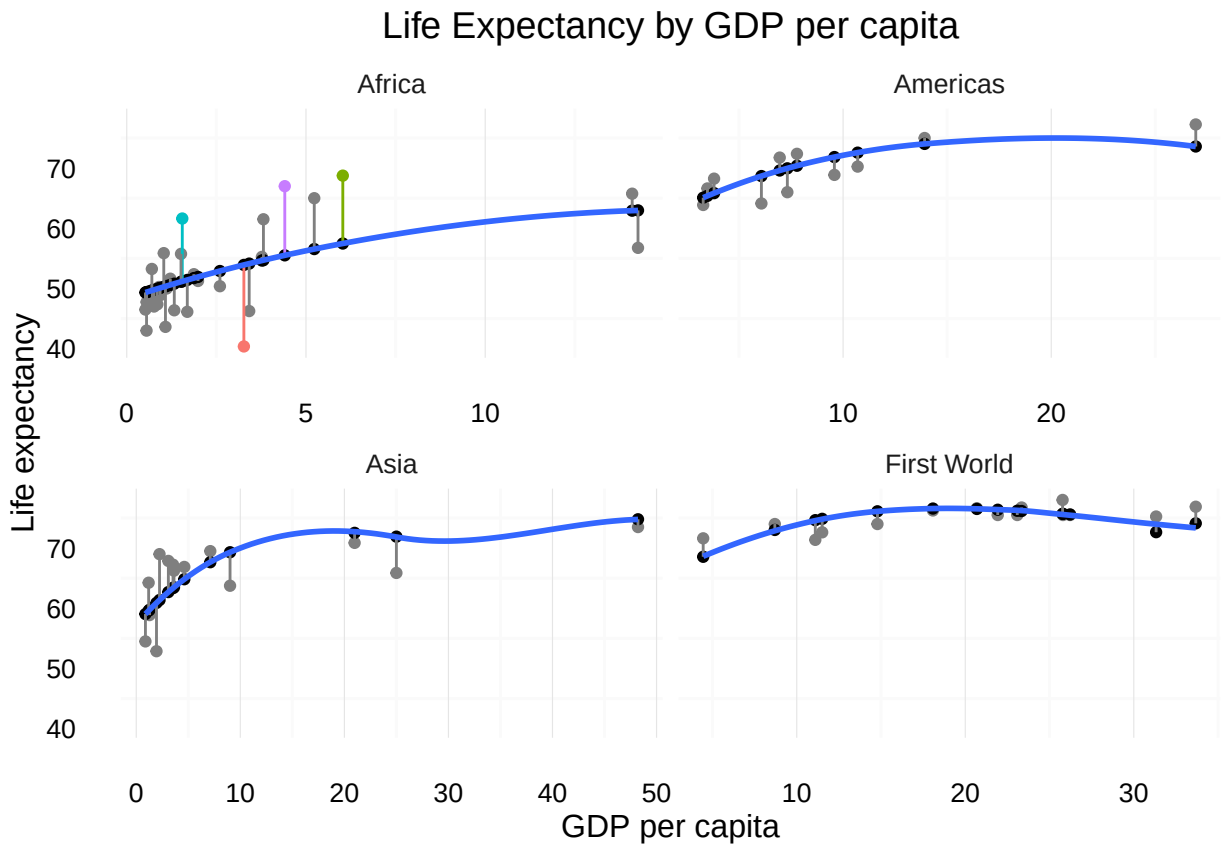
```
flash %>%
  select(continent_world,country,lifeExp,gdpPercap,popMill) %>%
  group_by(country,continent_world) %>%
  summarise_if(is.numeric,mean) %>%
  ungroup %>%
  mutate(predicted = predict(mod),
         difference = lifeExp-predicted,
         label = ifelse(abs(difference) > 9,as.character(country),NA)) -> flashForPlot
```

Table 3: First ten rows of flashForPlot

country	continent_world	lifeExp	gdpPercap	popMill	predicted	difference	label
Algeria	Africa	65.00	5.23	24.40	56.55	8.45	NA
Angola	Africa	40.38	3.27	8.61	53.92	-13.54	Angola
Australia	First World	76.75	24.00	16.84	76.75	0.00	NA
Austria	First World	75.50	25.80	7.82	75.69	-0.19	NA
Bahrain	Asia	70.88	20.98	0.48	72.54	-1.66	NA
Bangladesh	Asia	54.50	0.88	108.89	59.05	-4.55	NA
Bosnia and Herzegovina	First World	71.62	4.45	4.12	68.56	3.06	NA
Brazil	Americas	66.00	7.32	147.69	69.96	-3.96	NA
Burkina Faso	Africa	48.88	0.93	8.92	50.09	-1.21	NA
Burundi	Africa	46.50	0.52	5.55	49.33	-2.83	NA

The code below produces the plot below. The plot shows the difference between the fitted from the actual life expectancy value. The fitted values lie on the blue lines. The colored values represent the countries for which their average life expectancy is most different from what would be expected given their GDP per capita and continent.

```
flashForPlot %>%
  filter(country != "Australia") %>%
  ggplot() +
    geom_point(aes(gdpPercap, lifeExp,color=label)) +
    geom_point(aes(gdpPercap, predicted)) +
    geom_segment(aes(x = gdpPercap, y = lifeExp,
                     xend = gdpPercap, yend = predicted,
                     color = label)) +
    geom_smooth(aes(gdpPercap, predicted),se = FALSE) +
    facet_wrap( ~ continent_world, scales = "free_x") +
    theme_minimal() +
    theme(legend.position = "none",
          panel.grid.major.y = element_blank()) +
    xlab("GDP per capita") +
    ylab("Life expectancy") +
    ggtitle("Life Expectancy by GDP per capita")
```



Convergence

First, we create a vector of the names of the three statistics columns.

```
statNames <- names(select(flash,lifeExp:popMill))
```

```
## [1] "lifeExp"     "gdpPercap"   "gdpBillions" "popMill"
```

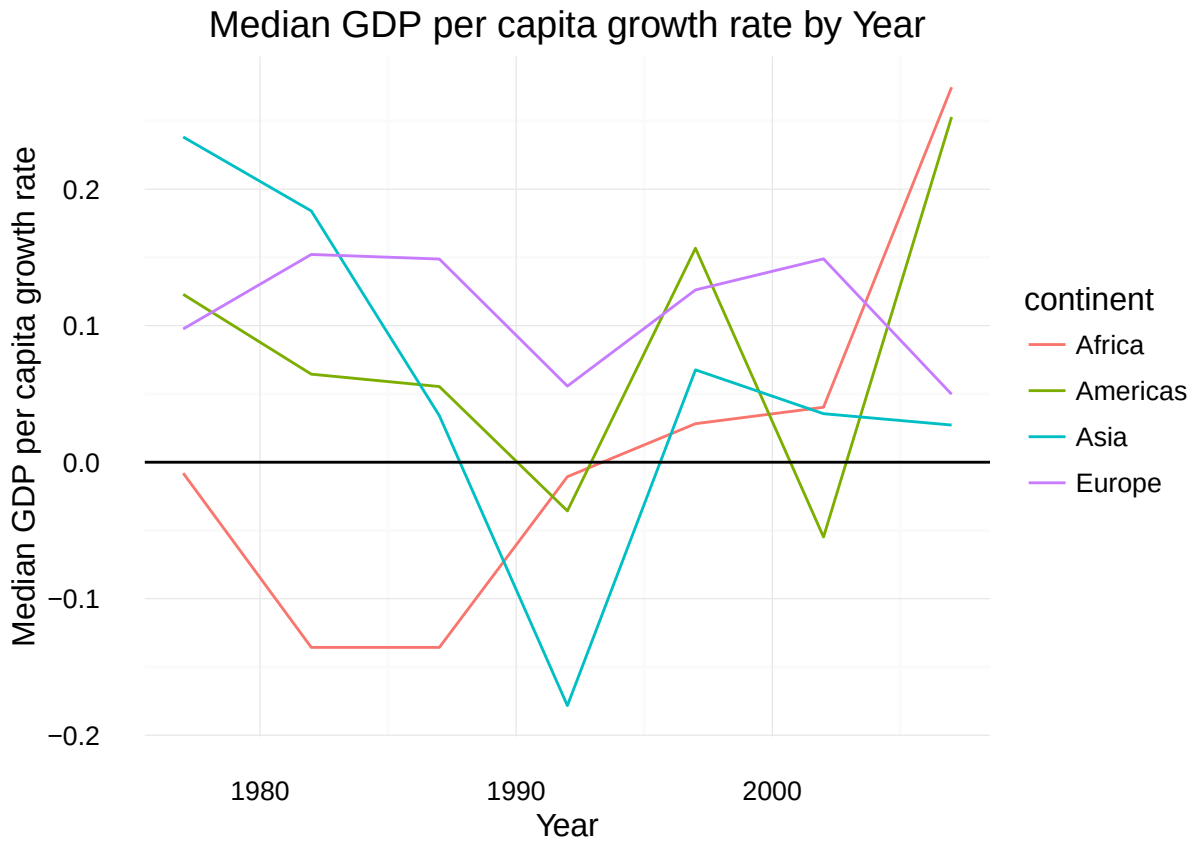
I use the lapply function to calculate growth rates.

```
continentList <- lapply(statNames,function(x){
  flash %>%
    spread_("continent",x) %>%
    group_by(year) %>%
    summarise_at(vars(Africa:Oceania),function(col) median(col,na.rm=T)) %>%
    select(-year) %>% #must remove year because we do not need its growth rate
    as.matrix %>% #growth rate transformation only possible as matrix
    log %>% diff %>% #calculate growth rates as the difference of the log values
    tbl_df %>% #mutate() only possible with tibble (or data frame)
    mutate(year = seq(1977,2007,5)) %>%
    melt("year",1:5,variable.name="continent",value.name=paste(x)) %>%
    tbl_df -> flash2
  if(which(statNames == x) == 1){flash2}#if in first loop, lapply returns three columns
  else{select_(flash2,x)}#if not in first loop, lapply returns newly calculated column
})
```

Using a for loop, I iteratively bind each successive column to the first data frame.

```
for(i in 2:4) continentList[[1]] <- bind_cols(continentList[[1]],continentList[i])
continentGrowthRates <- continentList[[1]]
```

```
continentGrowthRates %>%
  filter(continent != "Oceania") %>%
  ggplot(aes(year,gdpPercap,color=continent)) +
    geom_line() +
    geom_hline(yintercept = 0) +
    theme_minimal() +
    xlab("Year") +
    ylab("Median GDP per capita growth rate") +
    ggtitle("Median GDP per capita growth rate by Year")
```




```
    theme(legend.position = "bottom")
```

```
## List of 1
##  $ legend.position: chr "bottom"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```

Median GDP Per Capita

The chart below shows that first world countries have had a median GDP per capita that has been outpacing that of non first world countries in the Americas, Asia, and Africa.

```
flash %>%
  group_by(continent_world,year) %>%
  summarise(median_gdpPerCap = median(gdpPercap)) %>%
  ggplot(aes(year,median_gdpPerCap,color=continent_world)) +
    geom_line() +
    geom_hline(yintercept = 0) +
    theme_minimal() +
    xlab("Year") + ylab("Median GDP per capita") + ggtitle("Median GDP per capita by Year") +
    theme(legend.position = "bottom",
          legend.title = element_blank())
```