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

```r
flash <- tbl_df(read.csv("4_Flash_Proj_1_Data.csv"))

flash %>%
  mutate(country = str_replace_all(country,"[.,]",""),
         country = as.factor(country),
         gdpPercap = gdpPercap/1000,
         popMill = popThous/1000) %>%
  select(-popThous) -> flash
```

Table 1: First ten rows of manipulated flash data

country	continent	year	lifeExp	gdpPercap	gdpBillions	popMill
Algeria	Africa	1972	55	4.183	61.7	14.761
Algeria	Africa	1977	58	4.910	84.2	17.153
Algeria	Africa	1982	61	5.745	115.1	20.034
Algeria	Africa	1987	66	5.681	132.1	23.255
Algeria	Africa	1992	68	5.023	132.1	26.298
Algeria	Africa	1997	69	4.797	139.5	29.072
Algeria	Africa	2002	71	5.288	165.4	31.287
Algeria	Africa	2007	72	6.223	207.4	33.333
Angola	Africa	1972	38	5.473	32.3	5.895
Angola	Africa	1977	39	3.009	18.5	6.163

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)
(GDP/capita)^2	-18.399^{***}
	(5.805)
(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.

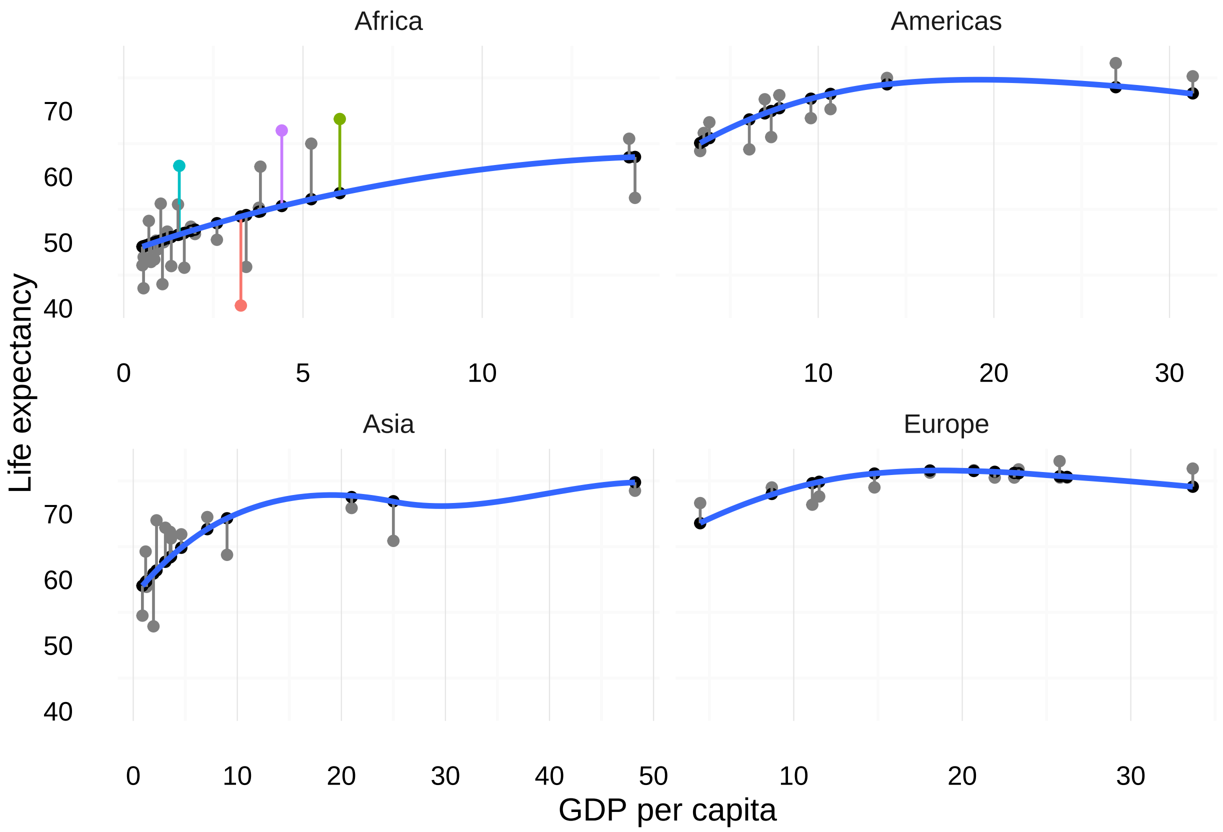
```r
flash %>%
  select(continent,country,lifeExp,gdpPercap,popMill) %>%
  group_by(country,continent) %>%
  summarise_if(is.numeric,mean) %>%
  ungroup %>%
  mutate(predicted = predict(mod),
         difference = lifeExp-predicted,
         label = ifelse(abs(difference) > 9,as.character(country),NA))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

Table 3: First ten rows of flashForPlot

country	continent	lifeExp	gdpPercap	popMill	predicted	difference	label
Algeria	Africa	65.000	5.231250	24.399125	56.54726	8.4527440	NA
Angola	Africa	40.375	3.268000	8.605625	53.91807	-13.5430679	Angola
Australia	Oceania	76.750	24.004375	16.840000	76.75000	0.0000000	NA
Austria	Europe	75.500	25.797250	7.824875	75.68555	-0.1855525	NA
Bahrain	Asia	70.875	20.984000	0.481750	72.53593	-1.6609348	NA
Bangladesh	Asia	54.500	0.882125	108.893750	59.05022	-4.5502220	NA
Bosnia and Herzegovina	Europe	71.625	4.450875	4.124625	68.56129	3.0637096	NA
Brazil	Americas	66.000	7.323625	147.687875	69.96282	-3.9628240	NA
Burkina Faso	Africa	48.875	0.931250	8.919250	50.08882	-1.2138214	NA
Burundi	Africa	46.500	0.521625	5.551625	49.33421	-2.8342098	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, scales = "free_x") +
    theme_minimal() +
    theme(legend.position = "none",
          panel.grid.major.y = element_blank()) +
    xlab("GDP per capita") +
    ylab("Life expectancy")
```



Convergence

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

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

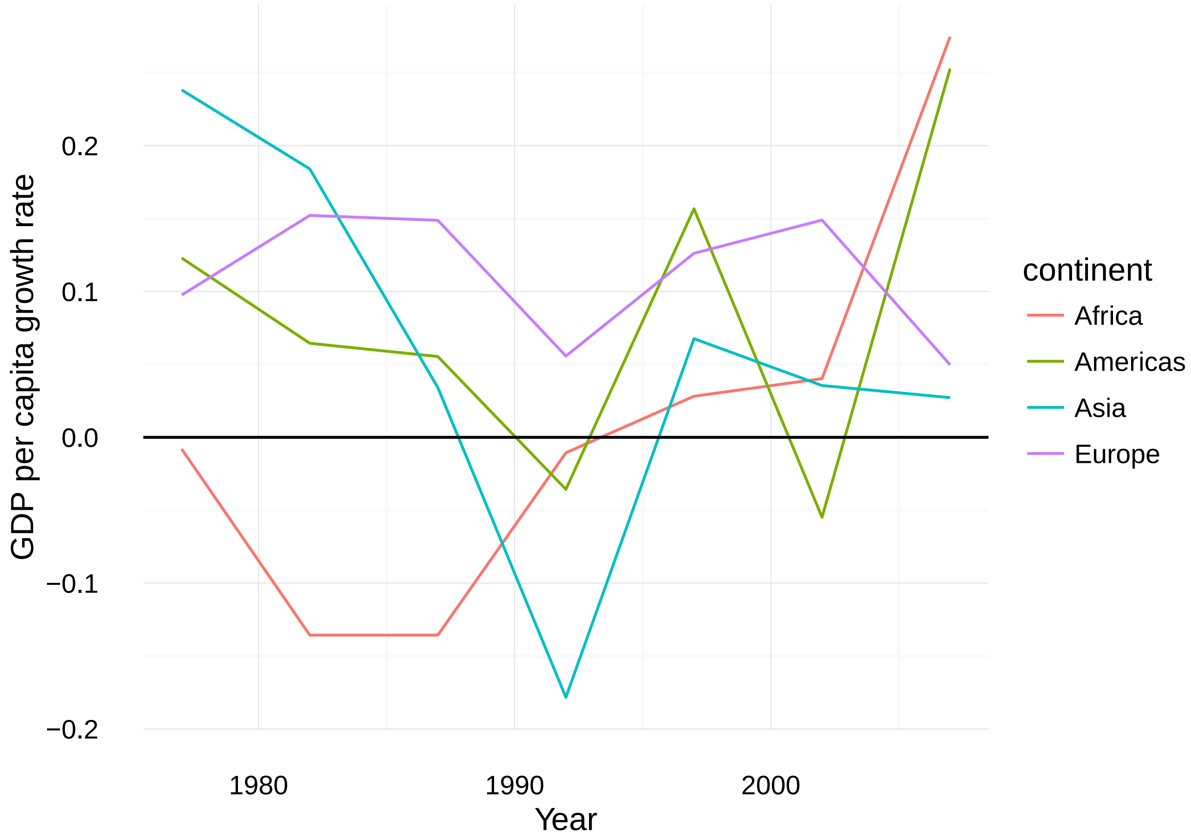
```
## [1] "lifeExp"     "gdpPercap"   "gdpBillions" "popMill"
```

I use the lapply function to calculate growth rates.

```r
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.

```r
for(i in 2:4) continentList[[1]] <- bind_cols(continentList[[1]],continentList[i])
continentGrowthRates <- continentList[[1]]
continentGrowthRates %>%
  filter(continent != "Oceania") %>%
  #gather(stat,growth_rate,lifeExp:popMill) %>%
  ggplot(aes(year,gdpPercap,color=continent)) +
  geom_line() +
  geom_hline(yintercept = 0) +
  #facet_wrap(~continent) +
  theme_minimal() +
  xlab("Year") +
  ylab("GDP per capita growth rate")
```


GDP per capita growth rate
Year
continent
Africa
Americas
Asia
Europe