

HW03: dplyr/ggplot2 Part II

Carleena Ortega

October 1 2019

Instructions

Pick three of the six tasks below, and produce:

- a tibble, using dplyr as your data manipulation tool
- an accompanying plot of data from the tibble, using ggplot2 as your visualization tool
- some dialogue about what your tables/figures show (doesn't have to be much).

Or, make up your own task(s)! If you want to do something comparable but different, i.e. swap one quantitative variable for another, be my guest! If you are feeling inspired and curious, then we're doing this right. Go for it. Just write down the task that you are now doing. Between the dplyr coverage in class and the list below, I think you get the idea.

Tasks

Task Option 1

Report the absolute and/or relative abundance of countries with low life expectancy over time by continent: Compute some measure of worldwide life expectancy – you decide – a mean or median or some other quantile or perhaps your current age. Then determine how many countries on each continent have a life expectancy less than this benchmark, for each year.

```
t1<-gapminder %>%
  select(continent,year,lifeExp) %>%
  group_by(year,continent) %>%
  filter(lifeExp<60) %>%
  group_by(year,continent) %>%
  summarize(low_LifeExp=length(continent))
knitr::kable(t1) %>%
  kable_styling("striped",latex_options="basic",full_width=FALSE,position="center")
```

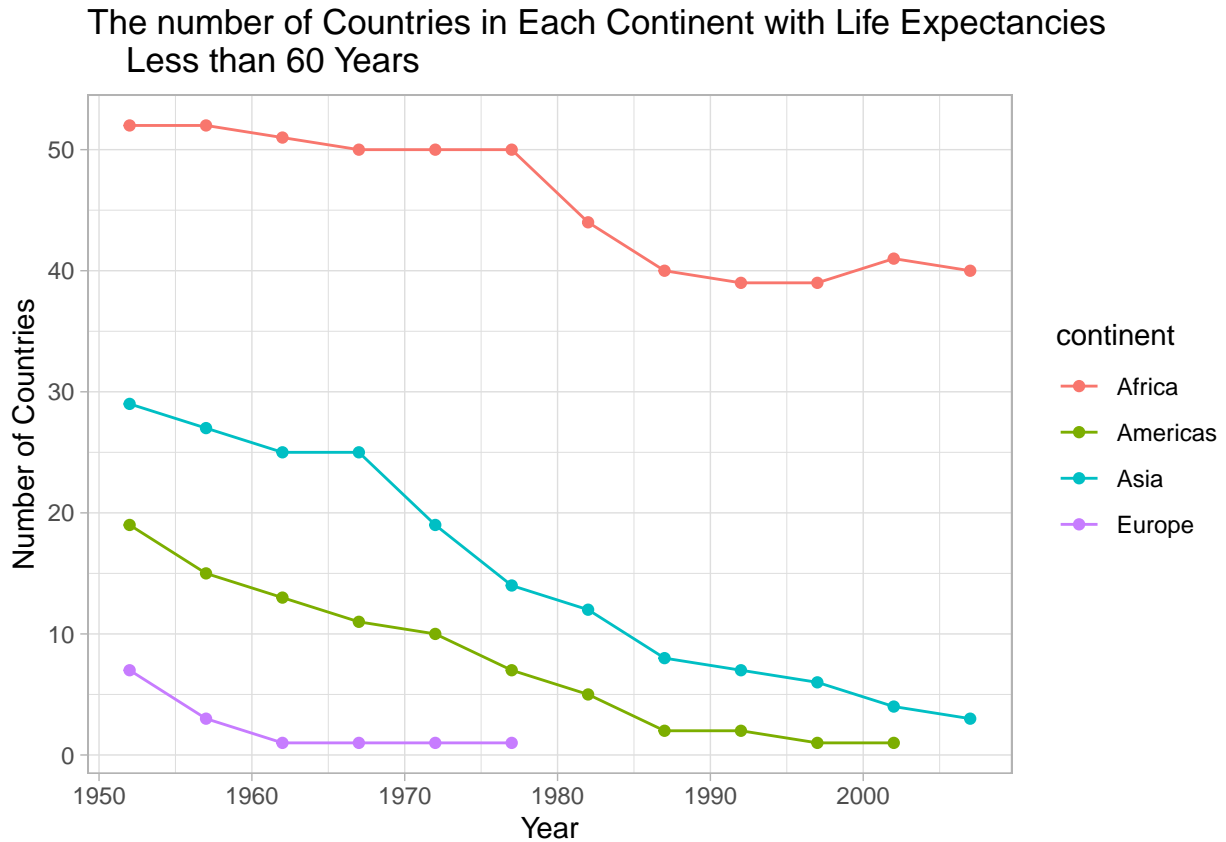
year	continent	low_LifeExp
1952	Africa	52
1952	Americas	19
1952	Asia	29
1952	Europe	7
1957	Africa	52
1957	Americas	15
1957	Asia	27
1957	Europe	3
1962	Africa	51
1962	Americas	13
1962	Asia	25
1962	Europe	1
1967	Africa	50
1967	Americas	11
1967	Asia	25
1967	Europe	1
1972	Africa	50
1972	Americas	10
1972	Asia	19
1972	Europe	1
1977	Africa	50
1977	Americas	7
1977	Asia	14
1977	Europe	1
1982	Africa	44
1982	Americas	5
1982	Asia	12
1987	Africa	40
1987	Americas	2
1987	Asia	8
1992	Africa	39
1992	Americas	2
1992	Asia	7
1997	Africa	39
1997	Americas	1
1997	Asia	6
2002	Africa	41
2002	Americas	1
2002	Asia	4
2007	Africa	40
2007	Asia	3

This table enumerates the number of countries that have life expectancies less than 60 years for each continent in every year recorded.

We can also present this data graphically as follows:

```
ggplot(t1)+
  geom_point(aes(year,low_LifeExp,group=continent,color=continent))+
  geom_path((aes(year,low_LifeExp,group=continent,color=continent)))+
  labs(x="Year",y="Number of Countries",
```

```
title="The number of Countries in Each Continent with Life Expectancies
Less than 60 Years")+
theme_light()
```



We can observe that the trends amongst the continents follow a similar path where the number of countries with life expectancies less than 60 decline throughout the years. Africa had a plateau from 1950 to late 1970's that declined sharply between late 1970 to late 1980 and then plateaued again until the late 2000's. Asia and Americas had a slow decline in the 1950's to the late 1960's then have a sharp decline from then on until it declines less rapidly from late 1980's to late 2000's. Europe had a sharp decline from 1950 to 1960 and experienced a plateau. We can attribute the decrease in the number of countries with life expectancies less than 60 as technology, sanitation, regulations, and healthcare improved throughout the years.

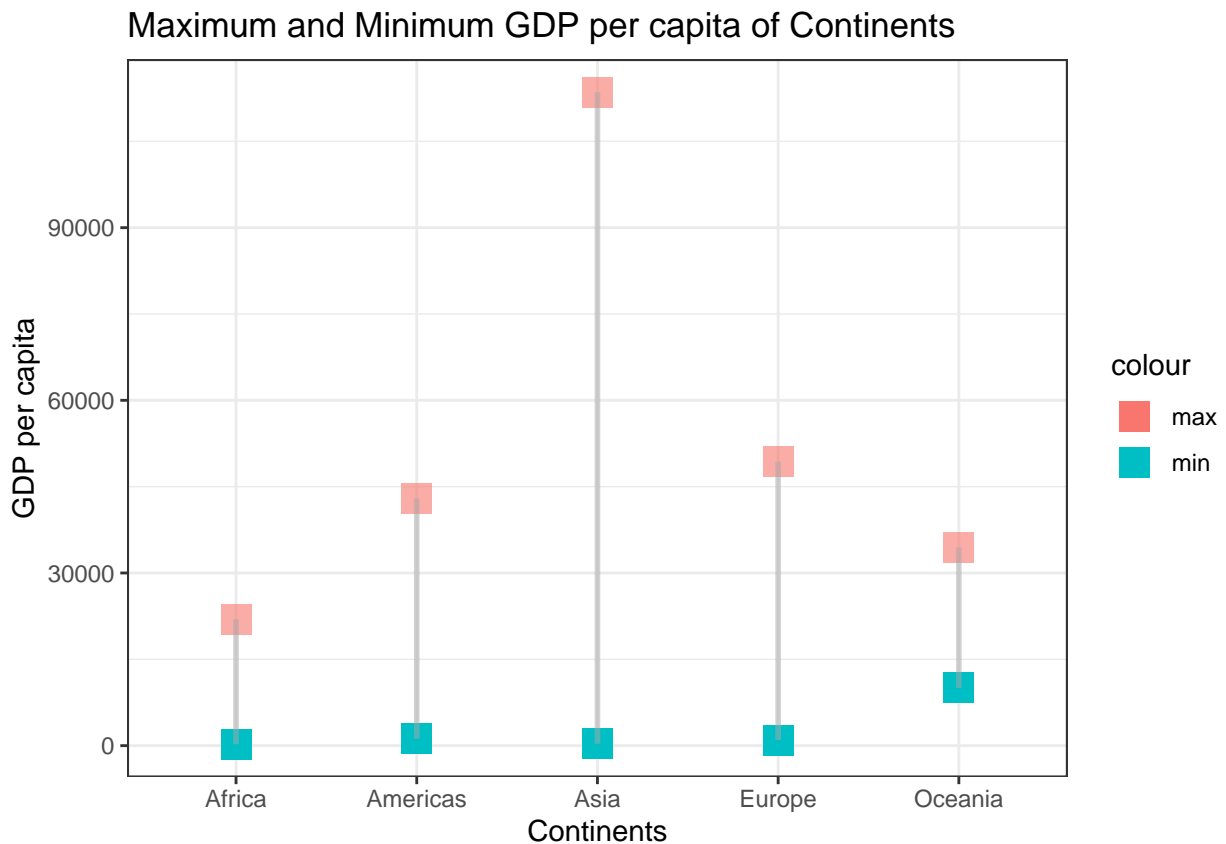
Task Option 2

Get the maximum and minimum of GDP per capita for all continents.

```
t2<-gapminder %>%
  select(continent,gdpPercap) %>%
  group_by(continent) %>%
  arrange(continent)%>%
  summarize(min_GDP=min(gdpPercap),max_GDP=max(gdpPercap))
knitr::kable(t2) %>%
  kable_styling("striped",latex_options="basic",full_width=FALSE,position="center")
```

continent	min_GDP	max_GDP
Africa	241.1659	21951.21
Americas	1201.6372	42951.65
Asia	331.0000	113523.13
Europe	973.5332	49357.19
Oceania	10039.5956	34435.37

```
ggplot(t2)+
  geom_point(aes(continent, min_GDP,color="min"),
    size=5, shape="square") + geom_point(aes(continent,max_GDP,colour="max"),
    size=5,shape="square",alpha=0.6)+
  geom_linerange(aes(x=continent,ymin=min_GDP,ymax=max_GDP),
    size=1,colour="dark gray",alpha=0.5)+
  labs(x="Continents",y="GDP per capita",
    title="Maximum and Minimum GDP per capita of Continents") +
  theme(legend.position="right")+
  theme_bw()
```



We can observe the minimum (pink box) and maximum (blue box) GDP per capita of each continent (connected by a gray line). Asia has the highest max GDP per capita while Africa has the lowest minimum GDP per capita.

Task Option 3

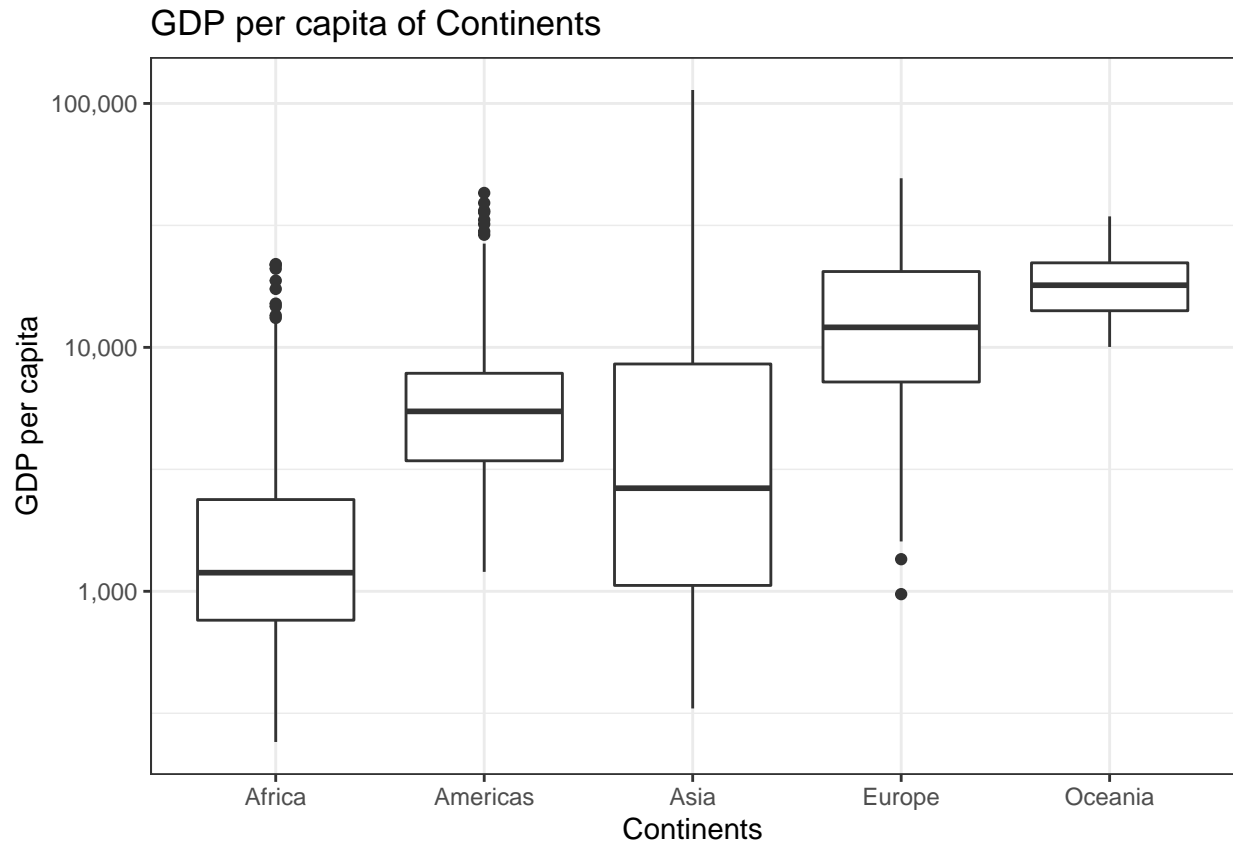
Look at the spread of GDP per capita within the continents.

```
gapminder %>%
  select(continent,gdpPercap) %>%
  group_by(continent) %>%
  summarize(mean=mean(gdpPercap),std_dev=sd(gdpPercap),
    median=median(gdpPercap),min=min(gdpPercap),max=max(gdpPercap)) %>%
  arrange(mean) %>%
knitr::kable() %>%
  kable_styling("striped",latex_options="basic",full_width=FALSE,position="center")
```

continent	mean	std_dev	median	min	max
Africa	2193.755	2827.930	1192.138	241.1659	21951.21
Americas	7136.110	6396.764	5465.510	1201.6372	42951.65
Asia	7902.150	14045.373	2646.787	331.0000	113523.13
Europe	14469.476	9355.213	12081.749	973.5332	49357.19
Oceania	18621.609	6358.983	17983.304	10039.5956	34435.37

This table provides a summary of the dataset. It lists the continents in increasing mean GDP per capita and also provides the standard deviation (std_dev), median, minimum, and maximum GDP per capita of each continent throughout the data collection.

```
gapminder %>%
  select(continent,gdpPercap) %>%
  group_by(continent) %>%
  arrange(continent)%>%
  ggplot()+
  geom_boxplot(aes(continent,gdpPercap)) +
  scale_y_log10(labels=comma)+
  labs(x="Continents",y="GDP per capita",
    title="GDP per capita of Continents")+
  theme_bw()
```



This is a boxplot of the GDP per capita of the various continents showing the minimum(lowest point of the vertical line), the maximum(highest point of the vertical line), and the median(horizontal line within the box) values for each continent's GDP per capita.

~~Task Option 4~~

Compute a trimmed mean of life expectancy for different years. Or a weighted mean, weighting by population. Just try something other than the plain vanilla mean.

~~Task Option 5~~

How is life expectancy changing over time on different continents?

~~Task Option 6~~

Find countries with interesting stories. Open-ended and, therefore, hard. Promising but unsuccessful attempts are encouraged. This will generate interesting questions to follow up on in class.