STAT 545A Assignment 03: dplyr/ggplot2 Part II

```
library(tidyverse)
library(ggplot2)
library(gapminder)
library(gridExtra)
library(grid)
knitr::opts_chunk$set(echo = TRUE)
```

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; and
- some dialogue about what your tables/figures show (doesn't have to be much).

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.

```
gapminder %>%
  select(country, year, continent, lifeExp) %>%
  mutate(n_low = ifelse(lifeExp < quantile(lifeExp,0.25), n(),NA)) %>%
  group_by(continent, year) %>%
  summarise(LE = mean(lifeExp),
            low = quantile(lifeExp,0.25),
            n_total = n(),
            n_low = mean(n_low)) #%>%
```

```
## # A tibble: 60 x 6
## # Groups:   continent [5]
##    continent  year    LE   low n_total n_low
##    <fct>     <int> <dbl> <dbl>   <int> <dbl>
##  1 Africa     1952  39.1  35.8      52    NA
##  2 Africa     1957  41.3  37.4      52    NA
##  3 Africa     1962  43.3  39.5      52    NA
##  4 Africa     1967  45.3  41.4      52    NA
##  5 Africa     1972  47.5  43.3      52    NA
##  6 Africa     1977  49.6  44.5      52    NA
##  7 Africa     1982  51.6  45.6      52    NA
##  8 Africa     1987  53.3  46.8      52    NA
##  9 Africa     1992  53.6  48.0      52    NA
## 10 Africa     1997  53.6  47.3      52    NA
## # ... with 50 more rows
```

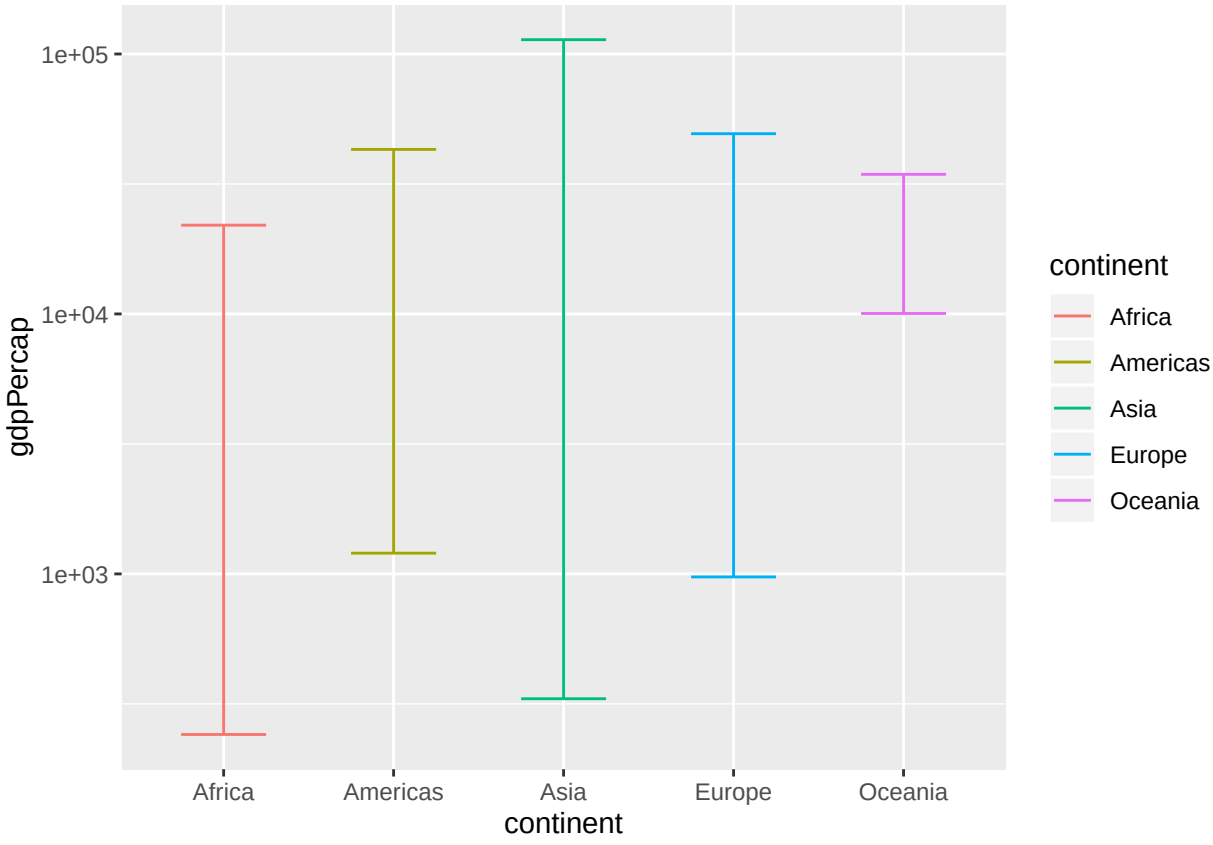
Task Option 2

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

```
rangeGdpP <- gapminder %>%
  group_by(continent) %>%
  summarize(min_gdpPercap = min(gdpPercap),
            max_gdpPercap = max(gdpPercap))
rangeGdpP
```

```
## # A tibble: 5 x 3
##   continent min_gdpPercap max_gdpPercap
##   <fct>             <dbl>         <dbl>
## 1 Africa              241.         21951.
## 2 Americas           1202.         42952.
## 3 Asia                331         113523.
## 4 Europe              974.         49357.
## 5 Oceania           10040.         34435.
```

```
ggplot(rangeGdpP) +
  geom_errorbar(aes(x=continent, ymin=min_gdpPercap, ymax=max_gdpPercap, color=continent, width=.5))  +
  scale_y_log10() +
  ylab("gdpPercap")
```



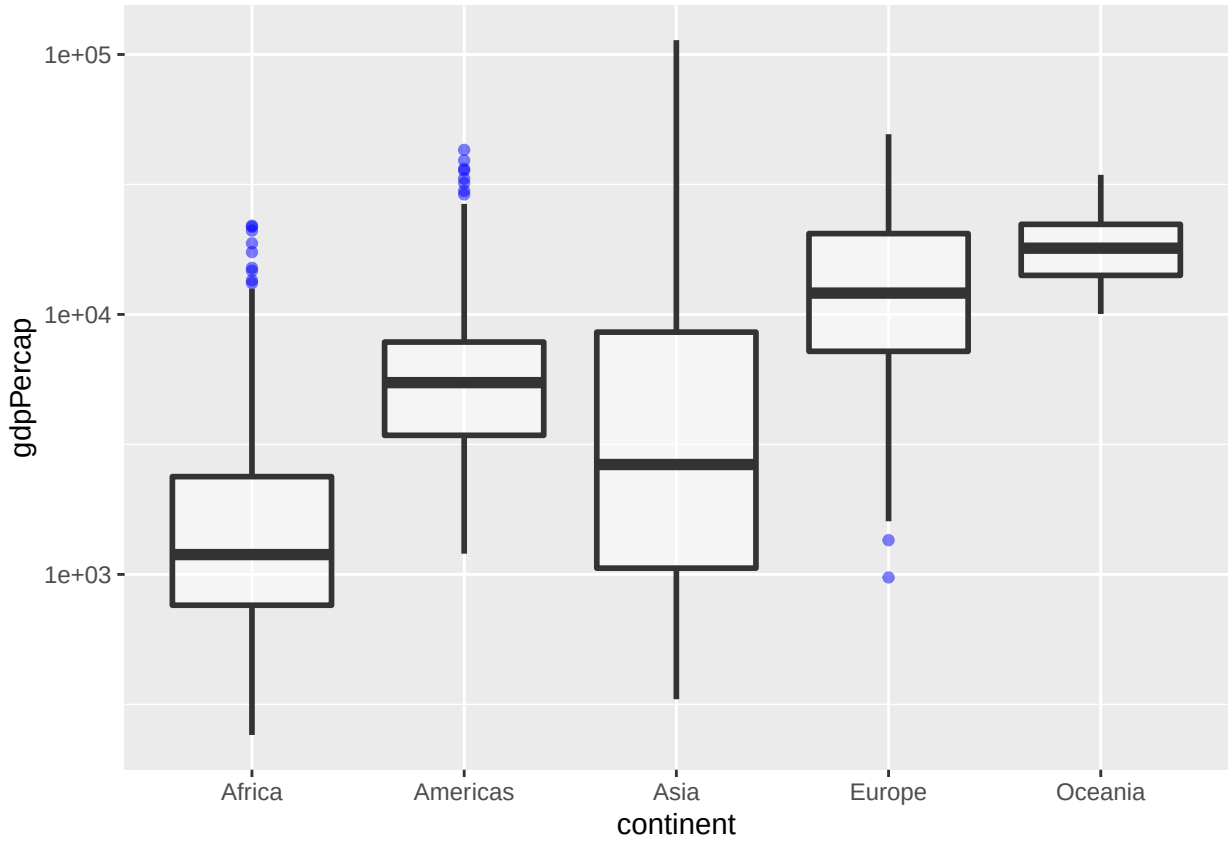
Task Option 3

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

```
gapminder %>%
group_by(continent) %>%
summarize(mean=mean(gdpPercap),
          min=min(gdpPercap),
          max=max(gdpPercap),
          std=sd(gdpPercap),
          q25=quantile(gdpPercap,0.25),
          q50=quantile(gdpPercap, 0.5),
          q75=quantile(gdpPercap, 0.75))
```

```
## # A tibble: 5 x 8
##   continent   mean    min     max    std    q25    q50    q75
##   <fct>      <dbl>  <dbl>   <dbl>  <dbl>  <dbl>  <dbl>  <dbl>
## 1 Africa     2194.   241.  21951.  2828.   761.  1192.  2377.
## 2 Americas   7136.  1202.  42952.  6397.  3428.  5466.  7830.
## 3 Asia       7902.   331  113523. 14045.  1057.  2647.  8549.
## 4 Europe    14469.   974.  49357.  9355.  7213. 12082. 20461.
## 5 Oceania   18622. 10040.  34435.  6359. 14142. 17983. 22214.
```

```
ggplot(gapminder, aes(x=continent, y=gdpPercap)) +
  geom_boxplot(outlier.colour = "blue", alpha=0.5, size=1, shape=1)  +
  scale_y_log10()
```



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

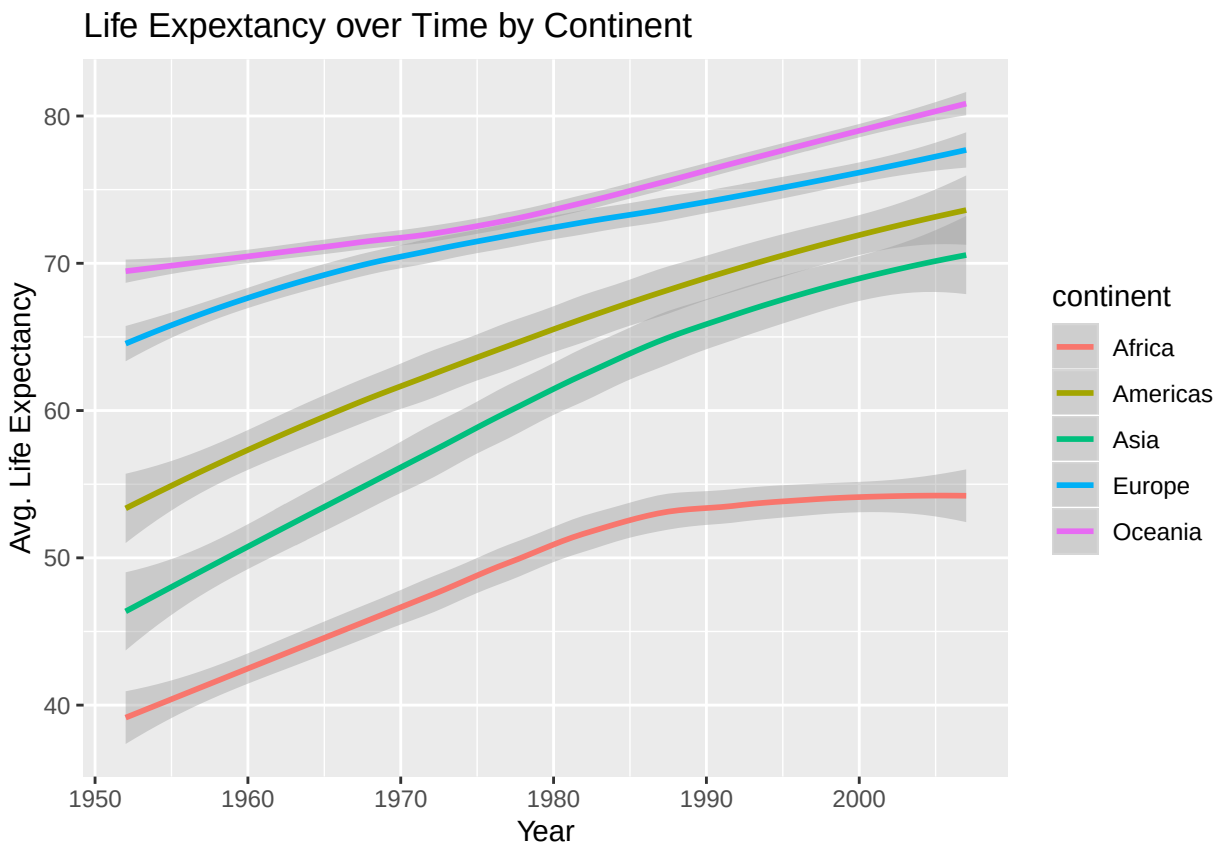
How is life expectancy changing over time on different continents?

```
gapminder %>%
    group_by(continent, year) %>%
    summarise(lifeExp=median(lifeExp))
```

```
## # A tibble: 60 x 3
## # Groups:   continent [5]
##    continent  year lifeExp
##    <fct>     <int>   <dbl>
##  1 Africa     1952    38.8
##  2 Africa     1957    40.6
##  3 Africa     1962    42.6
##  4 Africa     1967    44.7
##  5 Africa     1972    47.0
##  6 Africa     1977    49.3
##  7 Africa     1982    50.8
##  8 Africa     1987    51.6
##  9 Africa     1992    52.4
## 10 Africa     1997    52.8
## # ... with 50 more rows
```

```
ggplot(gapminder,
       aes(year,lifeExp, colour = continent)) +
  geom_smooth() + ggtitle("Life Expextancy over Time by Continent") +
  xlab("Year") + ylab("Avg. Life Expectancy")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



Task Option 6

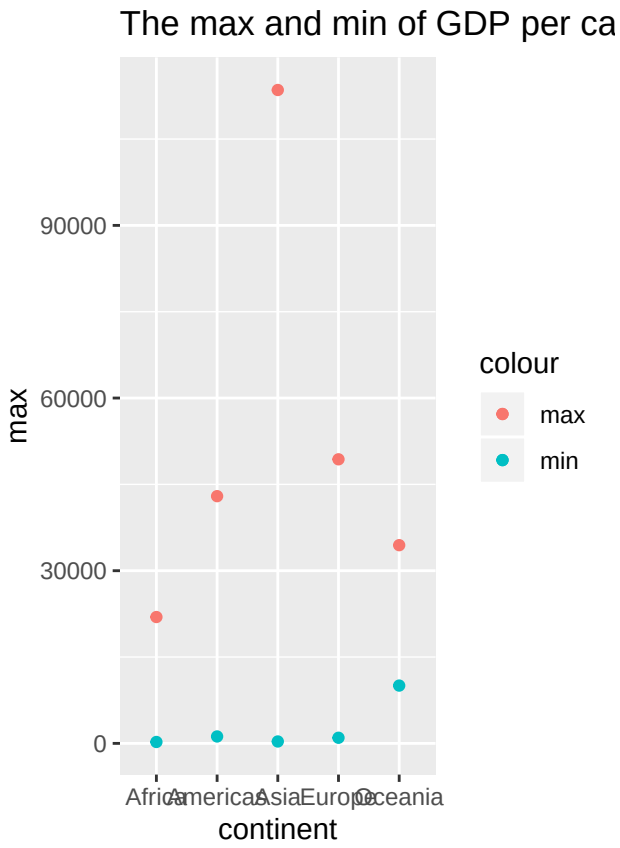
Find countries with interesting stories. Open-ended and, therefore, hard. Promising but unsuccessful attempts are encouraged. This will generate interesting questions to follow up on in class.

Ex

```
df <- gapminder %>%
  # for each continent
  group_by(continent) %>%
  summarize(max = max(gdpPercap), min = min(gdpPercap))
```

```
dfplt <- df %>%
  ggplot(aes(continent)) +
  # plot the points in the same graph and differentiate by color
  geom_point(aes(y = max, colour = "max")) +
  geom_point(aes(y = min, colour = "min")) +
  ggtitle("The max and min of GDP per capita for all continents")
# display the table and graph side by side
grid.arrange(tableGrob(df), dfplt, nrow = 1)
```

	continent	max	min
1	Africa	21951.21	241.1659
2	Americas	42951.65	1201.6372
3	Asia	113523.13	331.0000
4	Europe	49357.19	973.5332
5	Oceania	34435.37	10039.5956

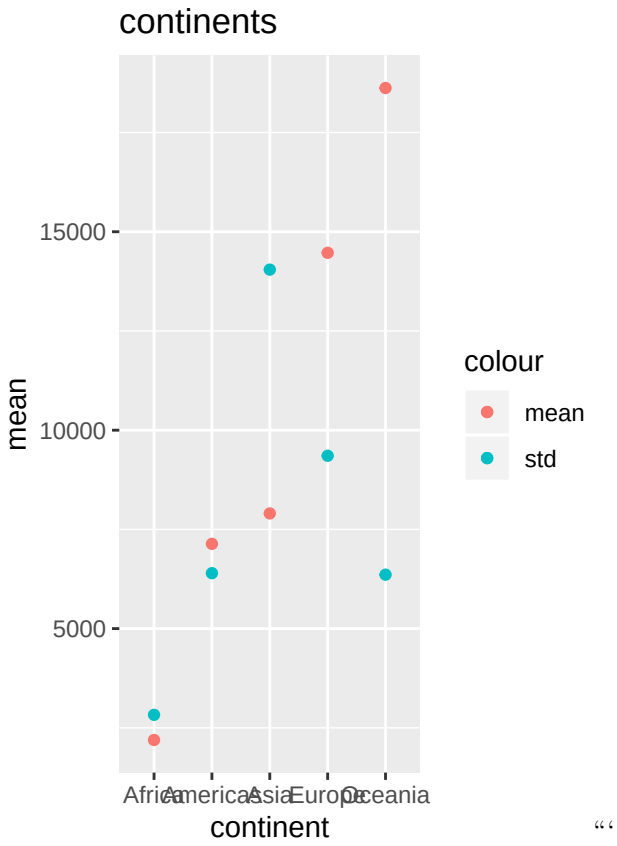



```
ex <- gapminder %>%
group_by(continent) %>%
summarize(mean=mean(gdpPercap),
          std=sd(gdpPercap))

exg <- ex %>%
  ggplot(aes(continent)) +
  geom_point(aes(y = mean, colour = "mean")) +
  geom_point(aes(y = std, colour = "std")) +
  ggtitle("continents")

  grid.arrange(tableGrob(ex), exg, nrow = 1)
```

	continent	mean	std
1	Africa	2193.755	2827.930
2	Americas	7136.110	6396.764
3	Asia	7902.150	14045.373
4	Europe	14469.476	9355.213
5	Oceania	18621.609	6358.983



“‘