Data Visulaization on Gapminder Dataframe

Vinodh Reddy Chennu
Reg n.o: 18MBMB09; MBA Business Analytics

Gapminder DataFrame

Gapminder data on life expectancy, GDP per capita and population by country.

The main data frame 'gapminder' has 1704 rows and 6 variables(columns).

Variables includes:

1. **country**: factor with 142 levels.
2. **continent**: factor with 5 levels.
3. **year**: ranges from 1952 to 2007 in increments of 5 years.
4. **lifeExp**: life expectancy at birth, in years.
5. **pop**: population.
6. **gdpPercap**: GDP per capita (US$, inflation-adjusted).

Summary of Gapminder Dataframe

```
summary(gapminder::gapminder)
```

```
##         country        continent        year         lifeExp
##  Afghanistan:  12   Africa  :624   Min.   :1952   Min.   :23.60
##  Albania    :  12   Americas:300   1st Qu.:1966   1st Qu.:48.20
##  Algeria    :  12   Asia    :396   Median :1980   Median :60.71
##  Angola     :  12   Europe  :360   Mean   :1980   Mean   :59.47
##  Argentina  :  12   Oceania : 24   3rd Qu.:1993   3rd Qu.:70.85
##  Australia  :  12                  Max.   :2007   Max.   :82.60
##  (Other)    :1632
##       pop              gdpPercap
##  Min.   :6.001e+04   Min.   :   241.2
##  1st Qu.:2.794e+06   1st Qu.:  1202.1
##  Median :7.024e+06   Median :  3531.8
##  Mean   :2.960e+07   Mean   :  7215.3
##  3rd Qu.:1.959e+07   3rd Qu.:  9325.5
##  Max.   :1.319e+09   Max.   :113523.1
##
```

Plotted Visualizations

- **Bar Chart** | TotalPop by Continents
- **Box Plot** | LifeExp by continents
- **ScatterPlot** | gdpPercap vs lifeExp
- **ScatterPlot** | gdpPercap vs lifeExp on log scale with continents differentiated with colors & pop with size
- **ScatterPlots with facetwrap** | gdpPercap vs LifeExp of each continent
- **Line graph** | Total population of continents over years
- **Line graph** | Population over the years and continents

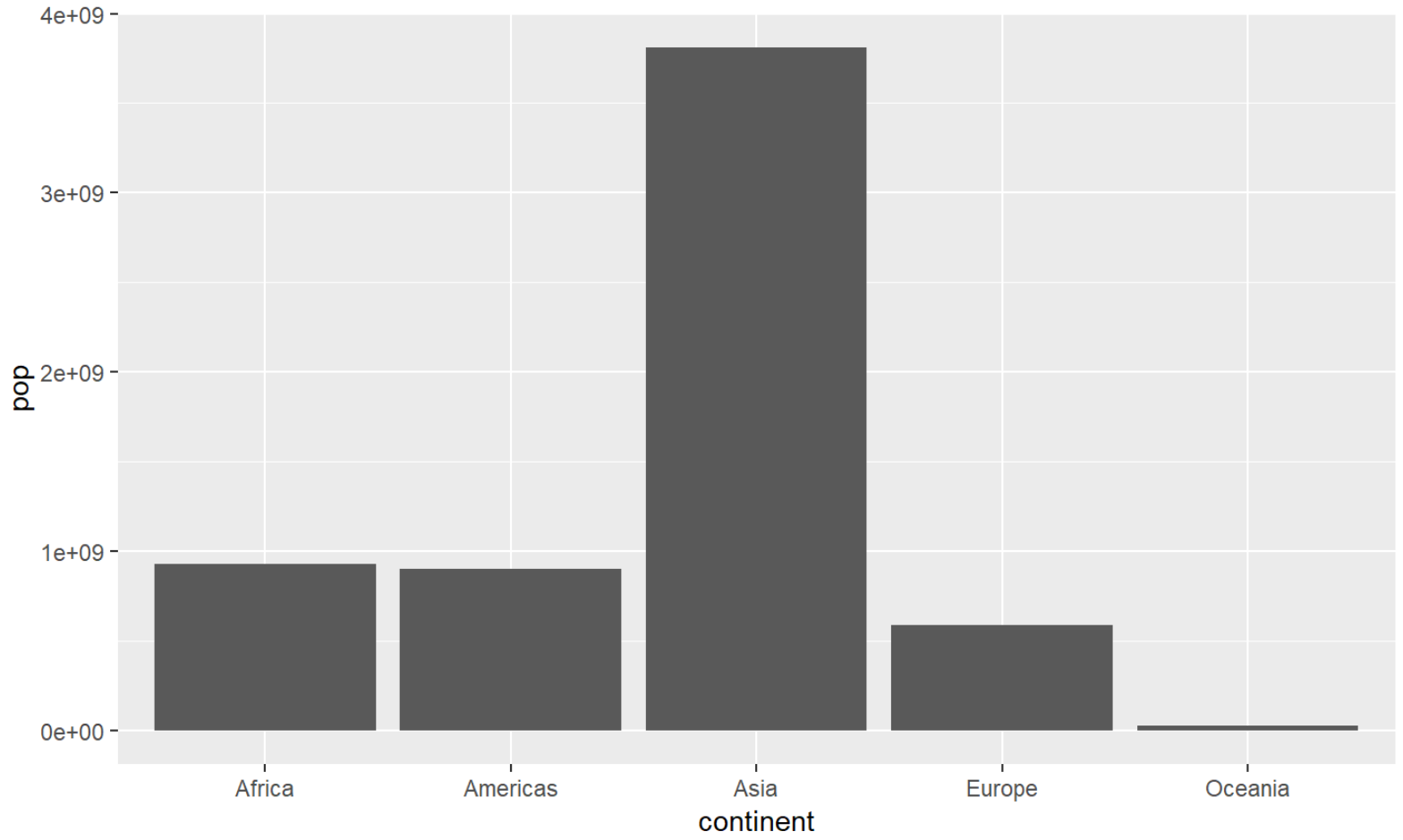
Create dataframe with filter year=2007

```
gapminder_year2007 <- gapminder::gapminder %>% filter(year == 2007)
gapminder_year2007
```

```
## # A tibble: 142 x 6
##    country     continent  year lifeExp       pop gdpPercap
##    <fct>       <fct>     <int>   <dbl>     <int>     <dbl>
##  1 Afghanistan Asia       2007    43.8  31889923      975.
##  2 Albania     Europe     2007    76.4   3600523     5937.
##  3 Algeria     Africa     2007    72.3  33333216     6223.
##  4 Angola      Africa     2007    42.7  12420476     4797.
##  5 Argentina   Americas   2007    75.3  40301927    12779.
##  6 Australia   Oceania    2007    81.2  20434176    34435.
##  7 Austria     Europe     2007    79.8   8199783    36126.
##  8 Bahrain     Asia       2007    75.6    708573    29796.
##  9 Bangladesh  Asia       2007    64.1 150448339     1391.
## 10 Belgium     Europe     2007    79.4  10392226    33693.
## # ... with 132 more rows
```

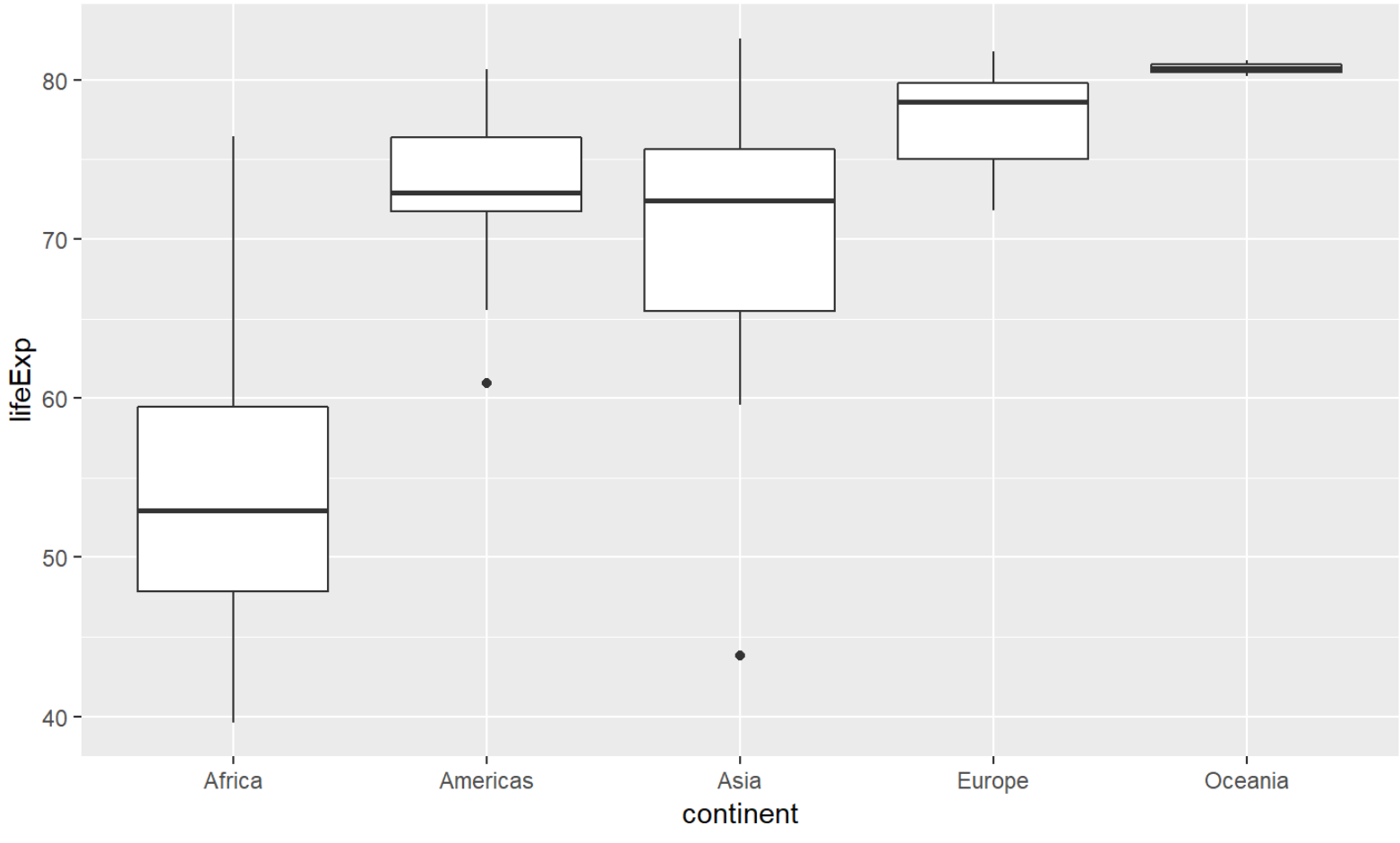
Population by continents

```
ggplot(gapminder_year2007, aes(x = continent, y = pop)) + geom_col()
```



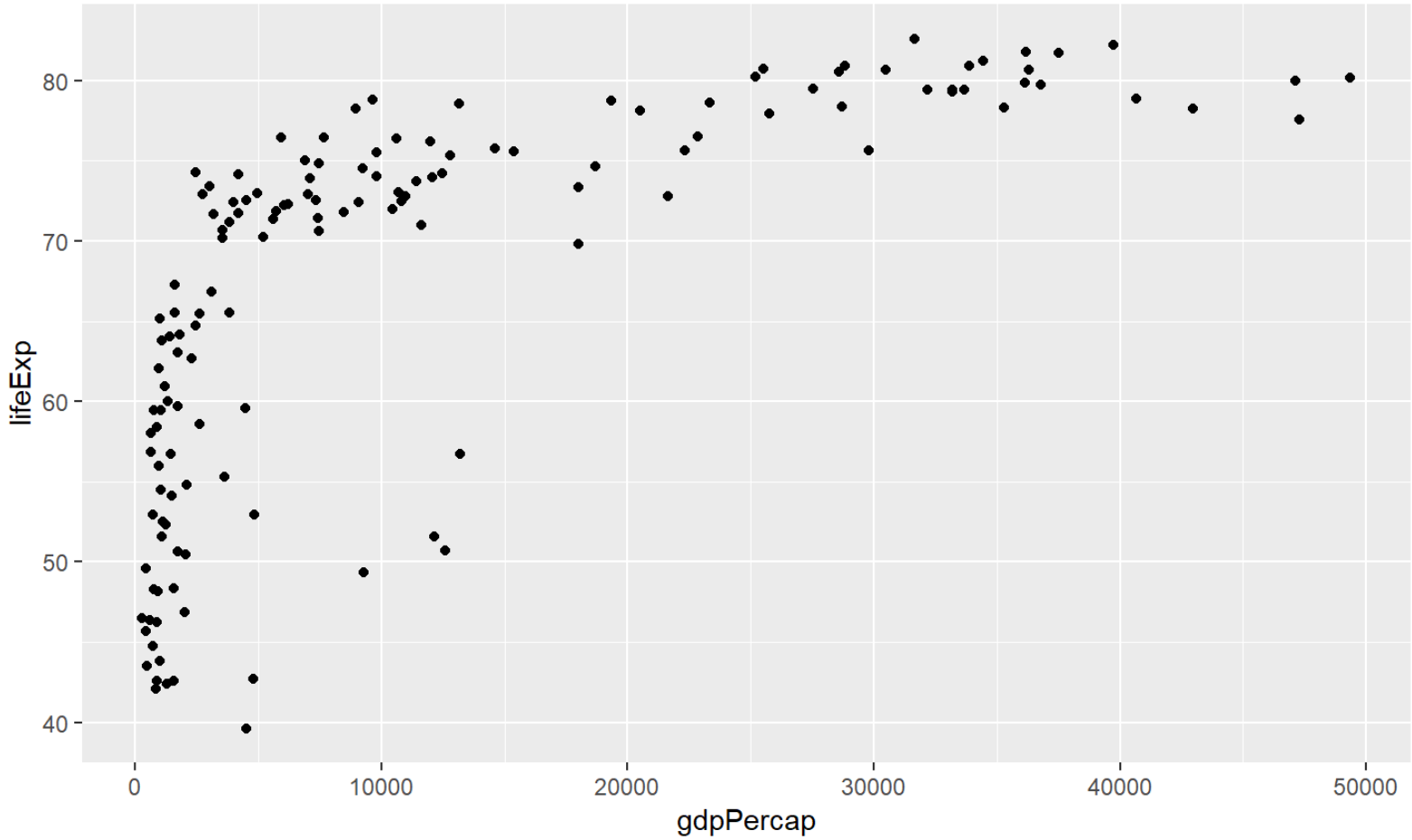
LifeExp by Continents

```
ggplot(gapminder_year2007, aes(x = continent, y = lifeExp)) + geom_boxplot()
```



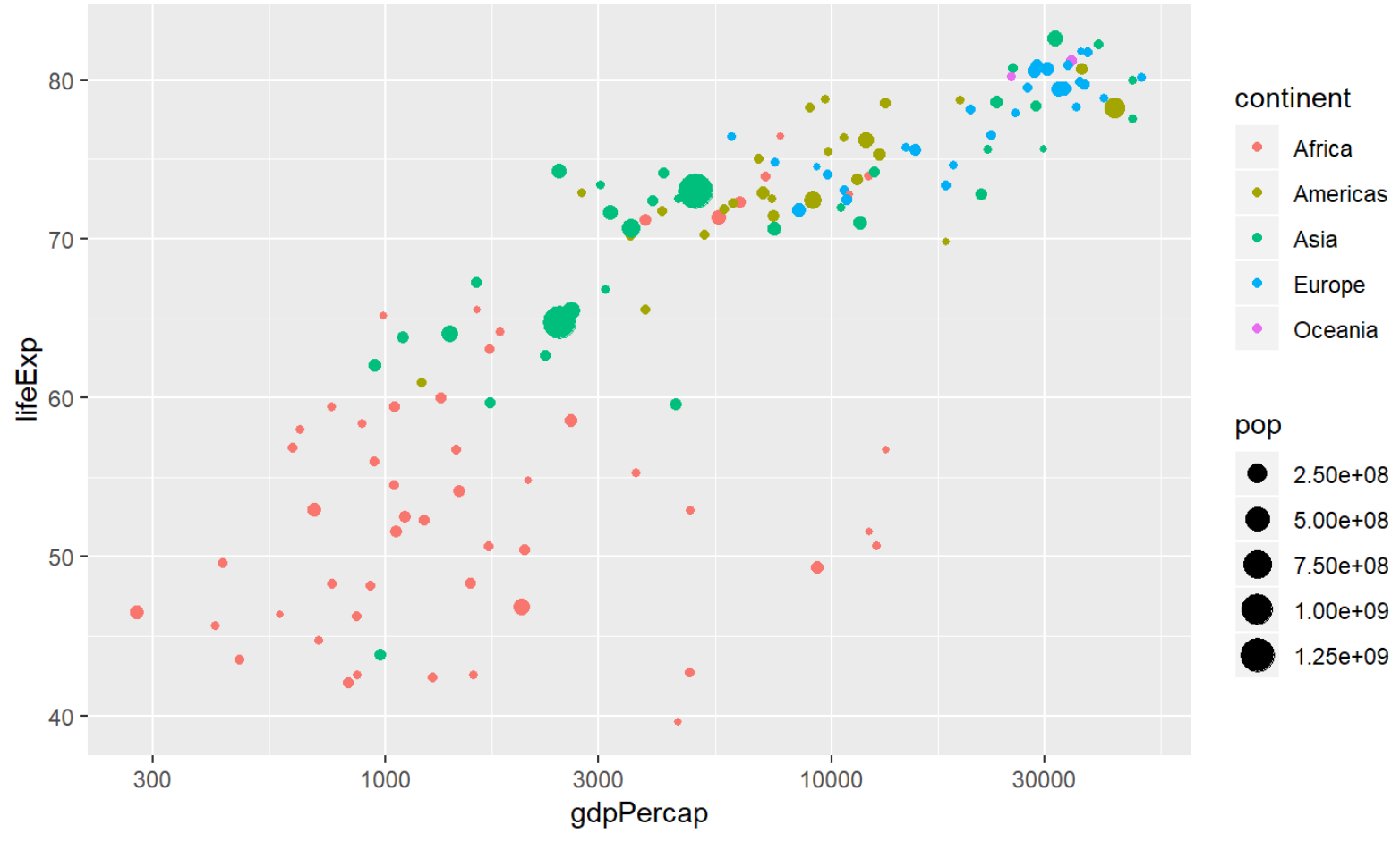
gdpPercap vs lifeExp

```
ggplot(gapminder_year2007, aes(x = gdpPercap, y = lifeExp)) + geom_point()
```



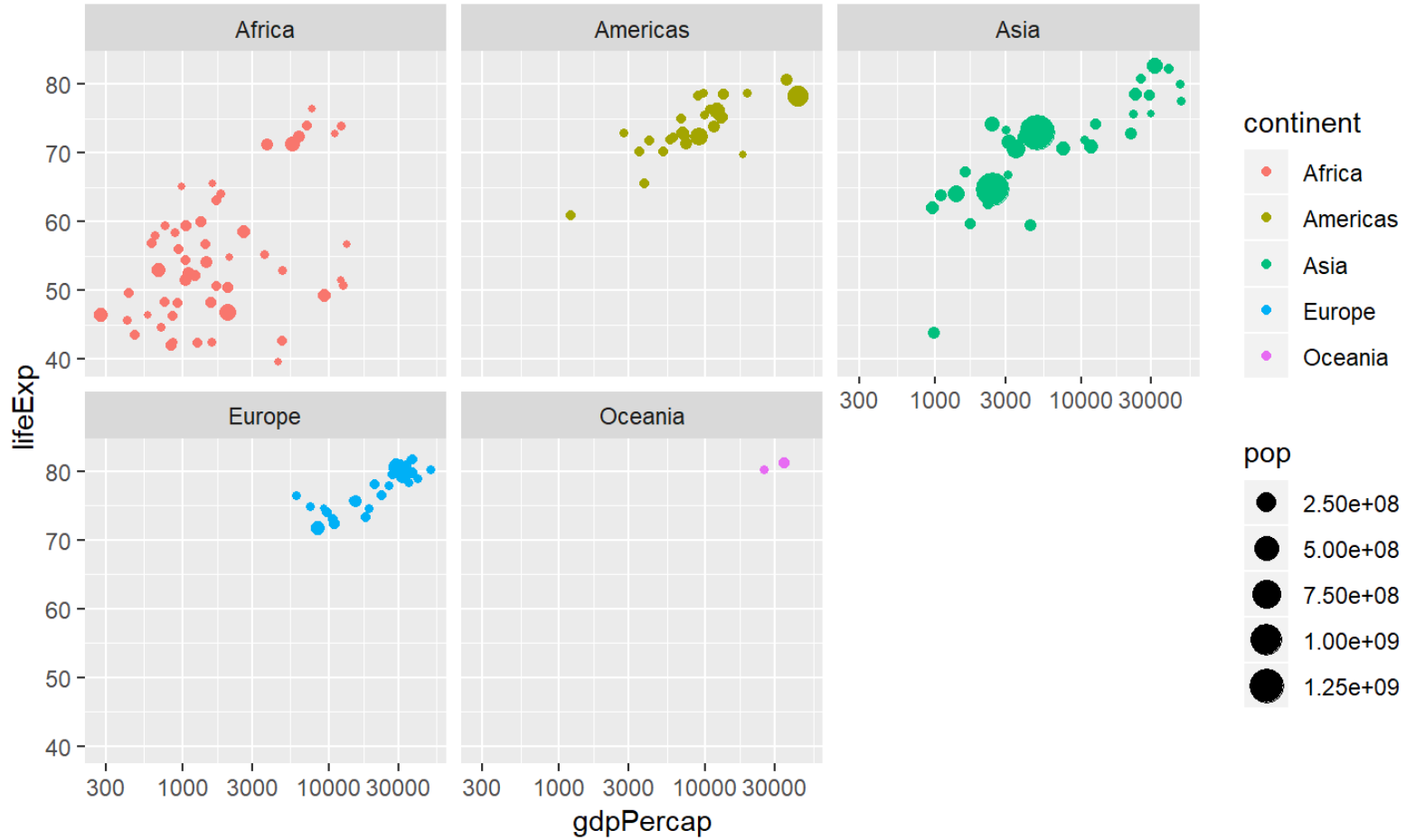
gdpPercap vs lifeExp with more options

```
ggplot(gapminder_year2007, aes(x = gdpPercap, y = lifeExp, color = continent, size=pop)) +
  geom_point() + scale_x_log10()
```



gdpPercap vs LifeExp of each continent

```
ggplot(gapminder_year2007, aes(x = gdpPercap, y = lifeExp,color = continent, size=pop)) +
  geom_point() + scale_x_log10() + facet_wrap(~ continent)
```



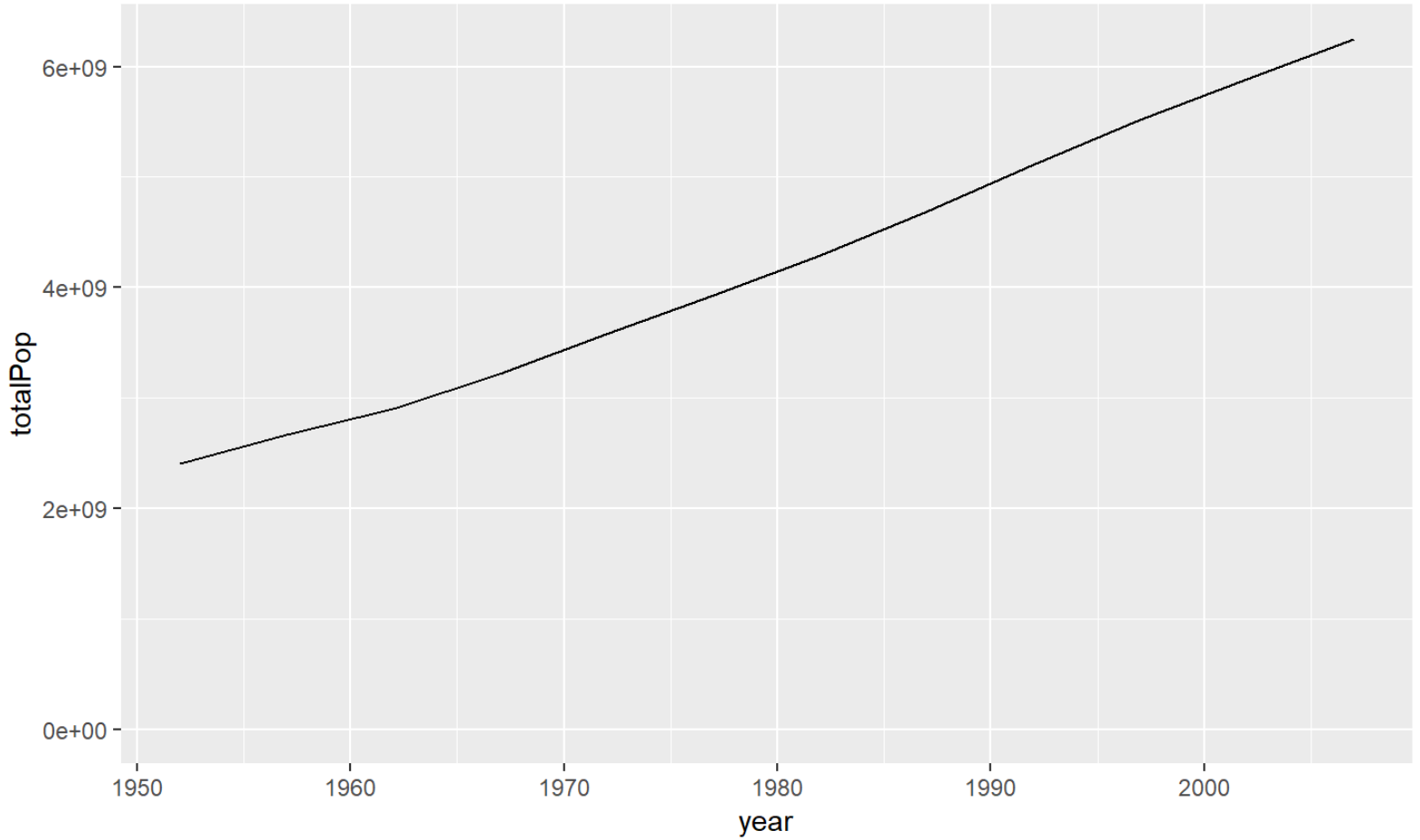
Created dataframe with groupby(year)

```
by_year <- gapminder::gapminder %>% group_by(year) %>% summarize(totalPop = sum(as.numeric(pop)),
 meanLifeExp = mean(lifeExp))
by_year
```

```
## # A tibble: 12 x 3
##     year   totalPop meanLifeExp
##    <int>      <dbl>       <dbl>
##  1  1952 2406957150        49.1
##  2  1957 2664404580        51.5
##  3  1962 2899782974        53.6
##  4  1967 3217478384        55.7
##  5  1972 3576977158        57.6
##  6  1977 3930045807        59.6
##  7  1982 4289436840        61.5
##  8  1987 4691477418        63.2
##  9  1992 5110710260        64.2
## 10  1997 5515204472        65.0
## 11  2002 5886977579        65.7
## 12  2007 6251013179        67.0
```

Total population over the years

```
ggplot(by_year, aes(x = year, y = totalPop)) + geom_line() + expand_limits(y = 0)
```



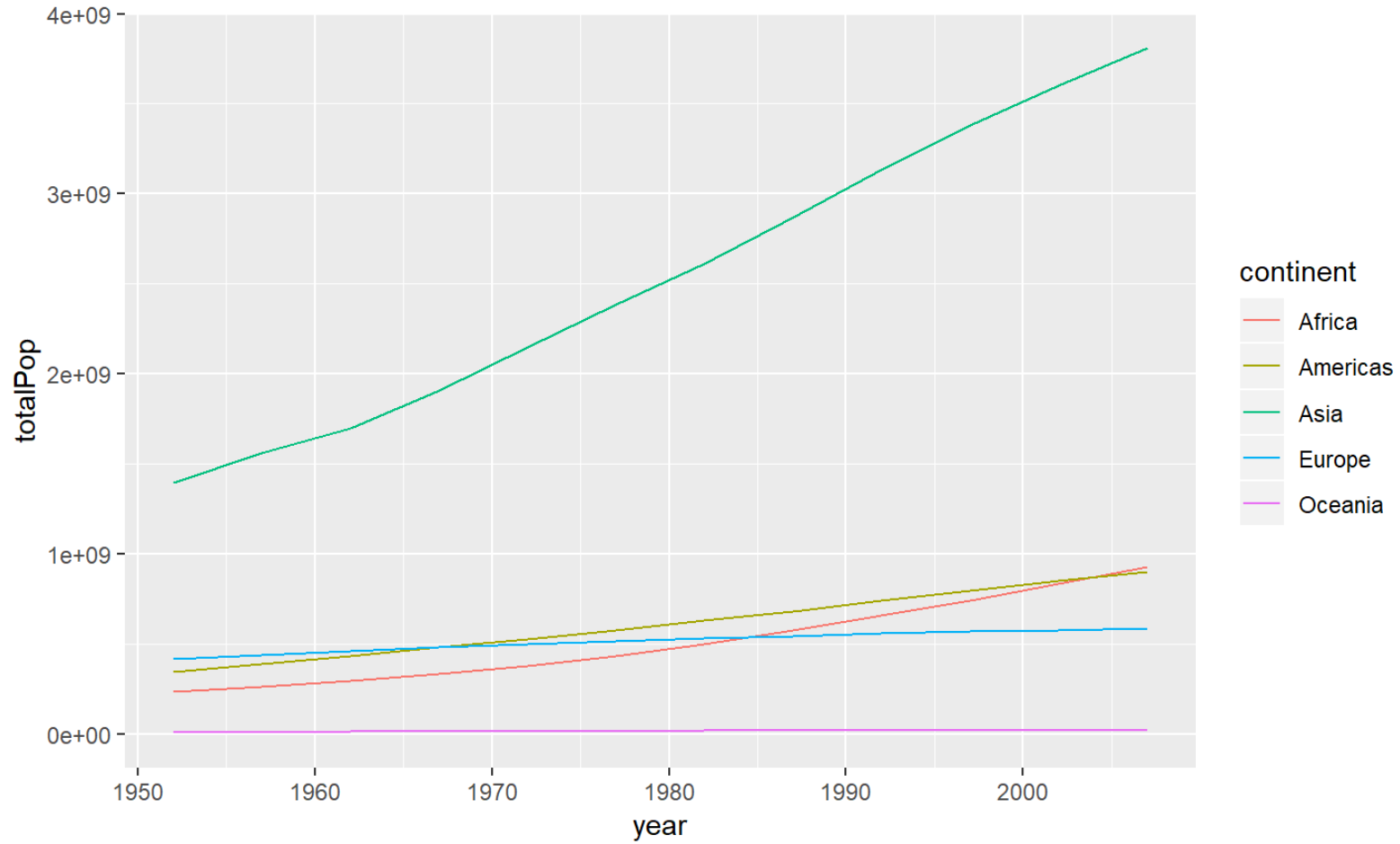
Data with groupby(year,continent)

```
by_year_continent <- gapminder::gapminder %>% group_by(year, continent) %>% summarize(totalPop
 = sum(as.numeric(pop)), meanLifeExp = mean(lifeExp))
by_year_continent
```

```
## # A tibble: 60 x 4
## # Groups:   year [?]
##     year continent   totalPop meanLifeExp
##    <int> <fct>          <dbl>       <dbl>
##  1  1952 Africa     237640501        39.1
##  2  1952 Americas   345152446        53.3
##  3  1952 Asia      1395357351        46.3
##  4  1952 Europe     418120846        64.4
##  5  1952 Oceania     10686006        69.3
##  6  1957 Africa     264837738        41.3
##  7  1957 Americas   386953916        56.0
##  8  1957 Asia      1562780599        49.3
##  9  1957 Europe     437890351        66.7
## 10  1957 Oceania     11941976        70.3
## # ... with 50 more rows
```

Population over the year by continents

```
ggplot(by_year_continent, aes(x = year, y = totalPop, color = continent)) + geom_line() + expand_limits(y = 0)
```



Conclusions

1. Asia continent has the higher population Whereas Ocenia continent has the least population.
2. Median lifeExp of Africa Continent is least and Ocenia continent is the highest.
3. On gdpPerCap vs lifeExp plots, most of the Africa countries lies at bottom left whereas Americas, Ocieania and European continents lie at the top right. There is a correlation between lifeExp and gdpPercap.
4. Total Population on the World increases over the time.
5. Total Population of Asia, Africa, Americas continents increases rapidly Whereas Europe and Ocenia changes very slightly over the time.