

Team 5

Yifan Feng: 2671027

Yunxiang Li: 2674844

1a)

```
> (out <- lm(formula = lifeExp ~ gdpPercap + pop + continent, data = gapminder))
> summary(out)

> summary(out)

Call:
lm(formula = lifeExp ~ gdpPercap + pop + continent, data = gapminder)

Residuals:
    Min       1Q   Median       3Q      Max
-49.161  -4.486   0.297   5.110  25.175

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.781e+01  3.395e-01 140.819 < 2e-16 ***
gdpPercap    4.495e-04  2.346e-05  19.158 < 2e-16 ***
pop          6.570e-09  1.975e-09   3.326 0.000901 ***
continentAmericas 1.348e+01  6.000e-01  22.458 < 2e-16 ***
continentAsia    8.193e+00  5.712e-01  14.342 < 2e-16 ***
continentEurope  1.747e+01  6.246e-01  27.973 < 2e-16 ***
continentOceania 1.808e+01  1.782e+00  10.146 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.365 on 1697 degrees of freedom
Multiple R-squared:  0.5821,    Adjusted R-squared:  0.5806
F-statistic: 393.9 on 6 and 1697 DF,  p-value: < 2.2e-16
```

1b)

```
> summary(gapminder)

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

```
gdpPercap
Min. : 241.2
1st Qu.: 1202.1
Median : 3531.8
Mean : 7215.3
3rd Qu.: 9325.5
Max. :113523.1
```

#summary(gapminder) gives us a summary of one dataset and represents selected data required in linear model. These data are min, max, Q1, Q3, mean and medium. All values are categorised in the variable: continent.

#summary(out) gives us the detailed statistics of the model, including additional coefficients, residual standard error, r-squared and F-stats.

1c)

#out\$coefficients demonstrates all coefficient values in each sector

#out\$fitted.values extracts fitted (expected and meaningful) values from one database. Values from the latter function should be distributed around the regression model.

2a)

#dim(pred_df) returns 6 columns (variables) and 500 rows.

2b)

#dim(pred_out) returns 3 columns (variables) and 500 rows.

2c)

#It contains two continent data (i.e. Africa and Europe) and excludes Americas, Asia, and Oceania.

2d)

#This graph utilizes `geom_point ()`, `geom_line()`, and `geom_ribbon()`.

2e)

```
> p <- ggplot(data = subset(pred_df, continent %in% c("Europe", "Africa")), aes(x =
gdpPercap, y = fit, ymin = lwr, ymax = upr, color = continent, fill = continent, group =
continent))
> p + geom_point(data = subset(gapminder,continent %in% c("Europe", "Africa")), aes(x =
gdpPercap, y = lifeExp, color = continent), alpha = 0.8, inherit.aes = FALSE) + geom_line()
```

```

+ labs(x = "GDP per Capita using 'scale_x_log10(labels = scales::dollar))", y = "Model
Fit") + theme_set(theme_bw()) + theme(legend.title=element_blank()) +
scale_x_log10(labels = scales::dollar) + geom_ribbon(alpha = 0.2, color = FALSE) +
scale_color_manual(values=c("darkseagreen4", "sienna2")) +
scale_fill_manual(values=c("darkseagreen4", "sienna2"))

```

