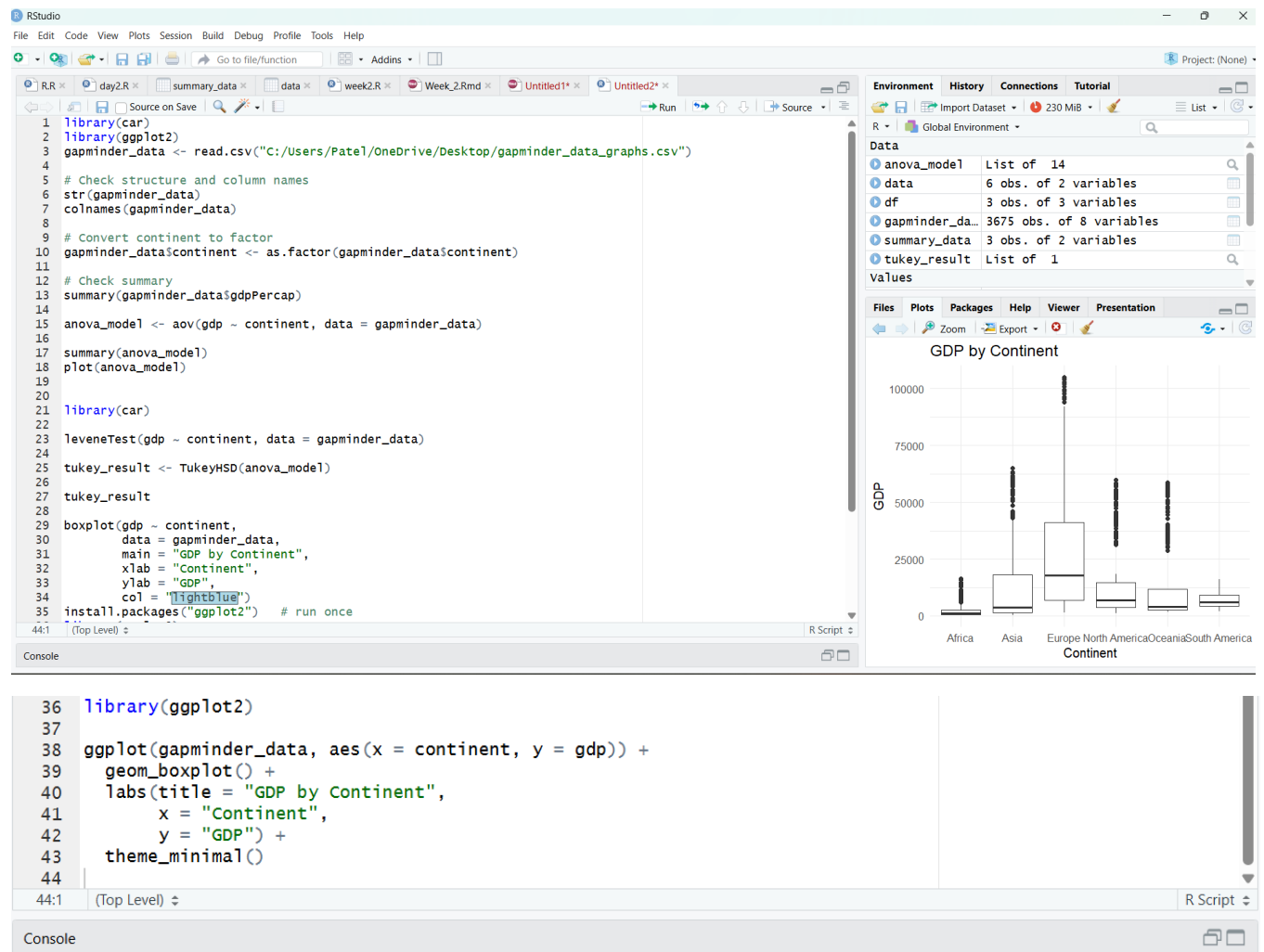
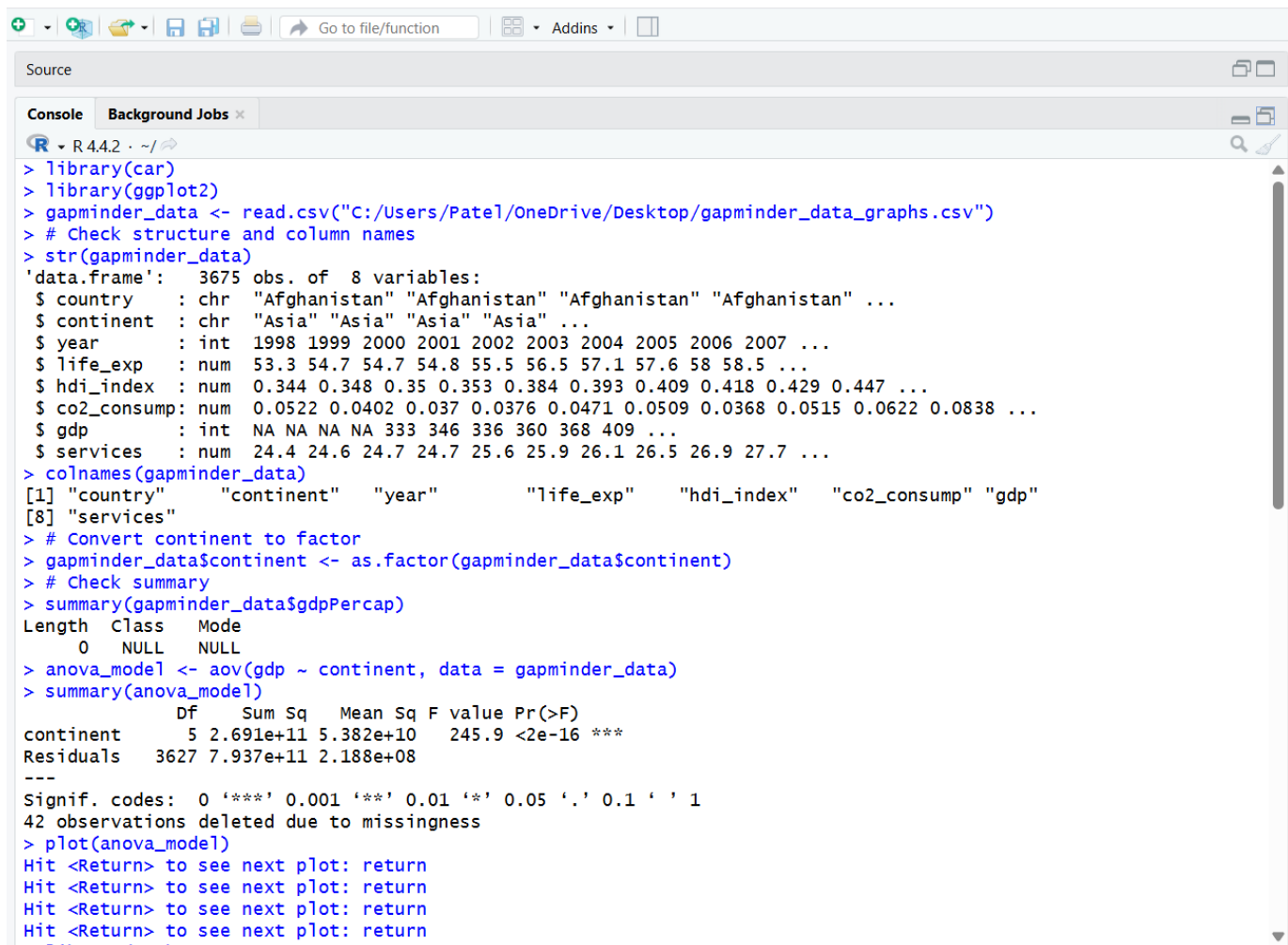


Code:-

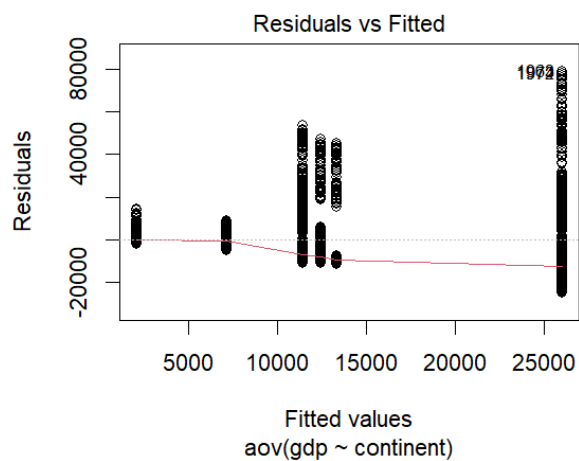
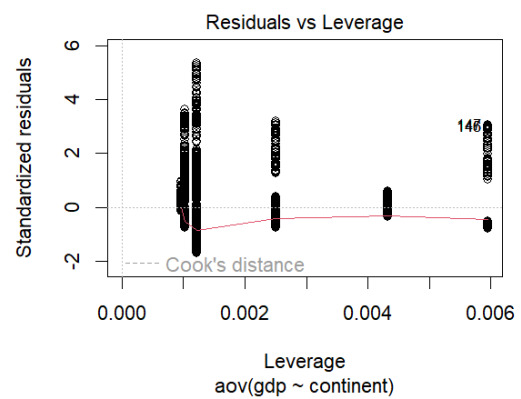
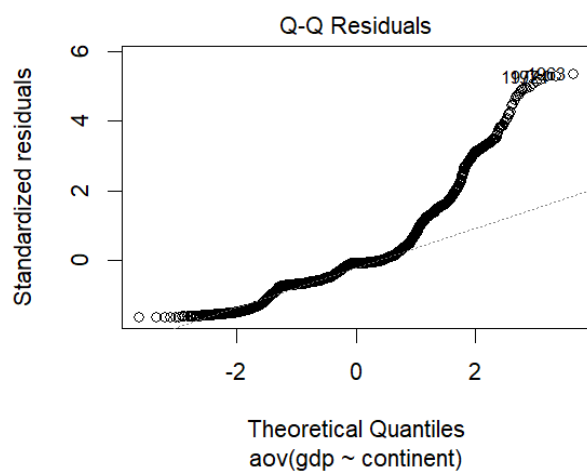
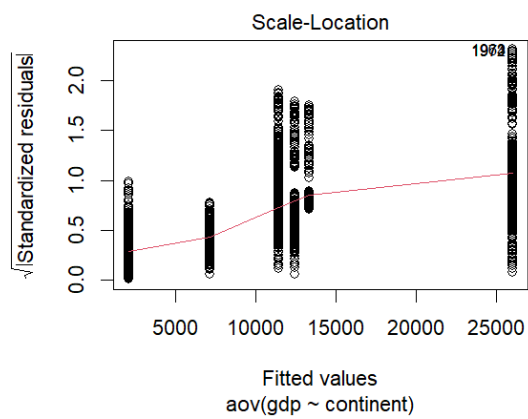
1. Use the Gapminder dataset.
2. Conduct ANOVA on GDP per capita by continent.
3. Check assumptions (e.g. using `plot(anova_model)`).
4. Run and interpret a Tukey post-hoc test.
5. Visualise the outcome.



Output:-



```
> library(car)
> library(ggplot2)
> gapminder_data <- read.csv("C:/Users/Patel/OneDrive/Desktop/gapminder_data_graphs.csv")
> # Check structure and column names
> str(gapminder_data)
'data.frame':   3675 obs. of  8 variables:
 $ country   : chr  "Afghanistan" "Afghanistan" "Afghanistan" "Afghanistan" ...
 $ continent : chr  "Asia" "Asia" "Asia" "Asia" ...
 $ year      : int   1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 ...
 $ life_exp  : num   53.3 54.7 54.7 54.8 55.5 56.5 57.1 57.6 58 58.5 ...
 $ hdi_index : num   0.344 0.348 0.35 0.353 0.384 0.393 0.409 0.418 0.429 0.447 ...
 $ co2_consump: num   0.0522 0.0402 0.037 0.0376 0.0471 0.0509 0.0368 0.0515 0.0622 0.0838 ...
 $ gdp       : int   NA NA NA 333 346 336 360 368 409 ...
 $ services  : num   24.4 24.6 24.7 24.7 25.6 25.9 26.1 26.5 26.9 27.7 ...
> colnames(gapminder_data)
[1] "country"    "continent"  "year"       "life_exp"   "hdi_index"  "co2_consump" "gdp"
[8] "services"
> # Convert continent to factor
> gapminder_data$continent <- as.factor(gapminder_data$continent)
> # Check summary
> summary(gapminder_data$gdpPercap)
Length Class Mode
0      NULL  NULL
> anova_model <- aov(gdp ~ continent, data = gapminder_data)
> summary(anova_model)
              Df Sum Sq Mean Sq F value Pr(>F)
continent      5 2.691e+11 5.382e+10  245.9 <2e-16 ***
Residuals    3627 7.937e+11 2.188e+08
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
42 observations deleted due to missingness
> plot(anova_model)
Hit <Return> to see next plot: return
Hit <Return> to see next plot: return
Hit <Return> to see next plot: return
Hit <Return> to see next plot: return
```



```

> library(car)
> leveneTest(gdp ~ continent, data = gapminder_data)
Levene's Test for Homogeneity of Variance (center = median)
      Df F value    Pr(>F)
group  5  183.58 < 2.2e-16 ***
3627
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> tukey_result <- TukeyHSD(anova_model)
> tukey_result
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = gdp ~ continent, data = gapminder_data)

$continent
      diff      lwr      upr    p adj
Asia-Africa    9355.4269   7475.916 11234.937 0.0000000
Europe-Africa  23937.5652  21965.879 25909.252 0.0000000
North America-Africa 10388.6584   7903.914 12873.402 0.0000000
Oceania-Africa  11288.7649   7781.006 14796.524 0.0000000
South America-Africa  5089.8796   2021.250   8158.509 0.0000344
Europe-Asia    14582.1383  12584.845 16579.431 0.0000000
North America-Asia  1033.2314  -1471.880   3538.343 0.8484258
Oceania-Asia     1933.3379  -1588.878   5455.554 0.6217907
South America-Asia -4265.5474  -7350.692  -1180.402 0.0011571
North America-Europe -13548.9069 -16123.896 -10973.917 0.0000000
Oceania-Europe  -12648.8004 -16221.053  -9076.547 0.0000000
South America-Europe -18847.6856 -21989.836 -15705.535 0.0000000
Oceania-North America    900.1065  -2978.991   4779.204 0.9860205
South America-North America -5298.7788  -8785.828  -1811.730 0.0002191
South America-Oceania  -6198.8853 -10475.558  -1922.212 0.0005238

> library(ggplot2)
> ggplot(gapminder_data, aes(x = continent, y = gdp)) +
+   geom_boxplot() +
+   labs(title = "GDP by Continent",
+         x = "Continent",
+         y = "GDP") +
+   theme_minimal()

```

