2016 Olympics: Medal Analysis
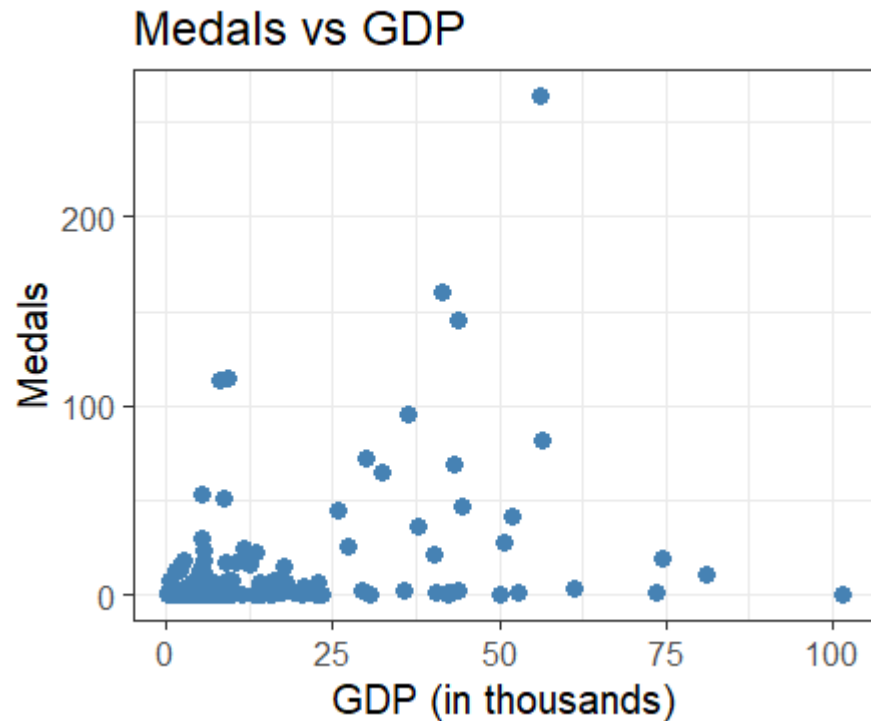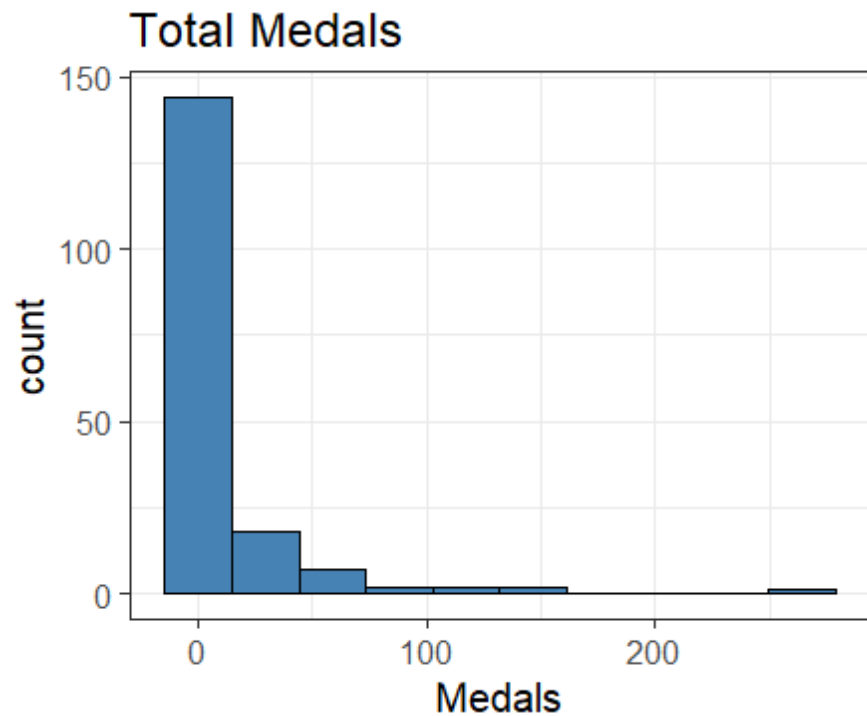
By: Jessica Murphy, Alex Rotondo, and Alex Schwartz

Research Question

Which of the following factors are associated with the number of medals a country earned in the 2016 Olympics?

- GDP per capita
- Number of athletes per population
- Median age of athletes
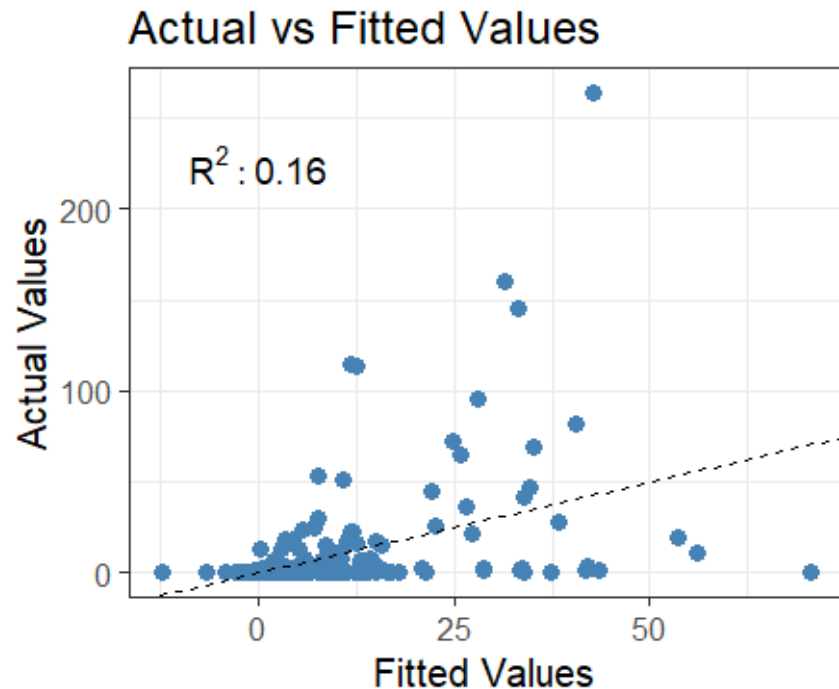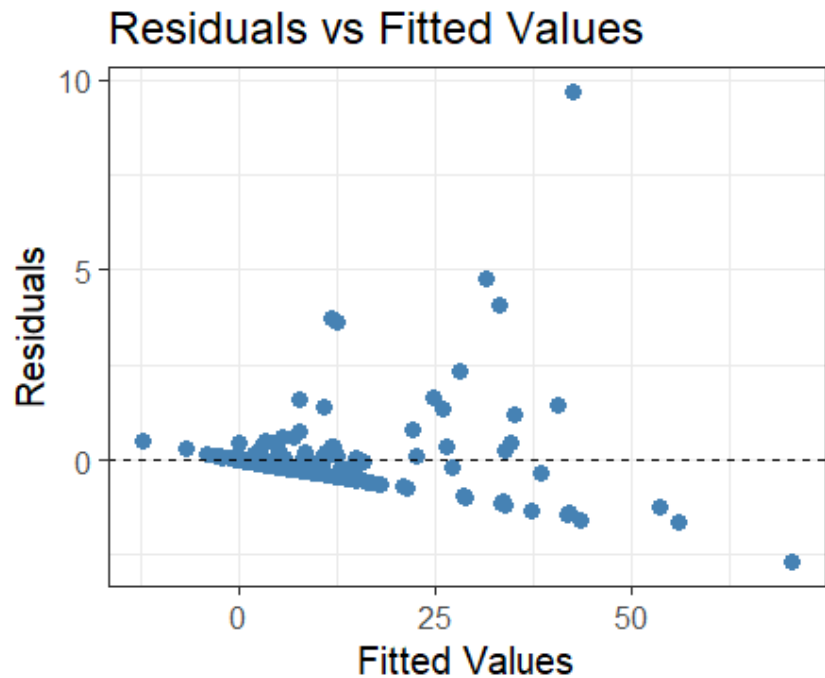- Percentage of female athletes

Exploratory Plots

Linear Model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \; gdp + \hat{\beta}_2 \; athletes + \hat{\beta}_3 \; age + \hat{\beta}_4 \; females$$

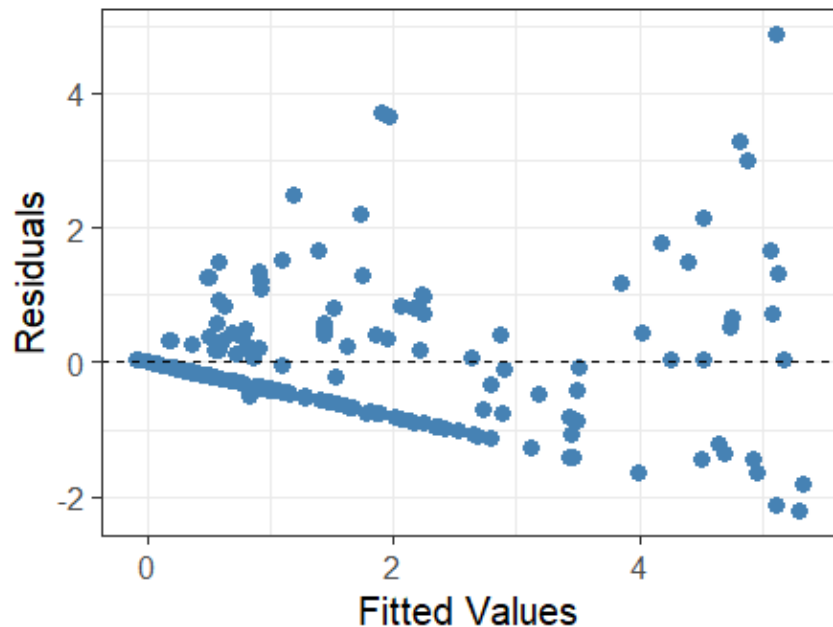|	Coefficient	Std. Error	t value	P-value
(Intercept)	-16.82	22.27	-0.76	0.45
gdp	0.64	0.13	4.96	1.7e-6
athletes	-0.74	0.79	-0.94	0.35
age	0.51	0.88	0.59	0.56
females	0.19	0.15	1.30	0.20

Diagnostic Plots

Variable Selection / Transformations

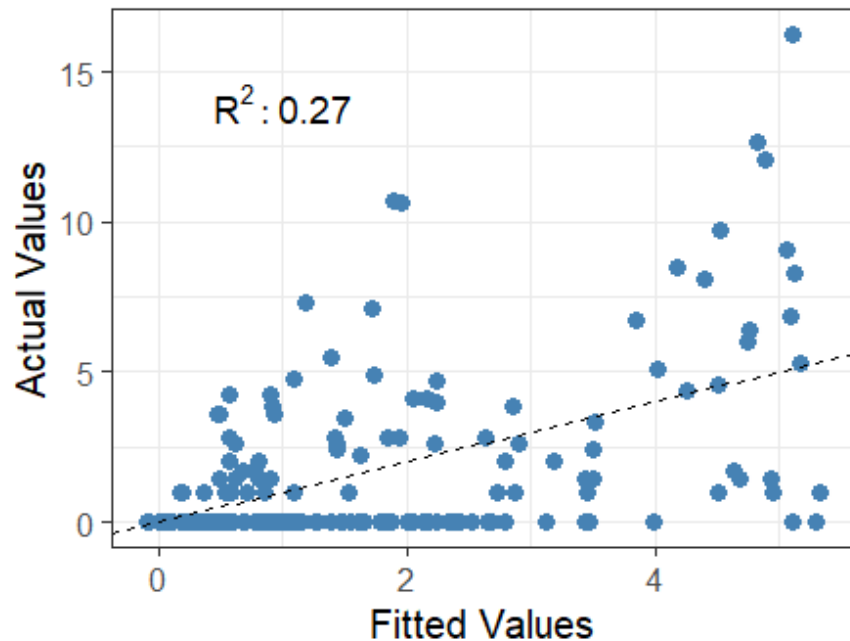$$\sqrt{\hat{Y}} = \hat{\beta}_0 + \hat{\beta}_1 \, gdp + \hat{\beta}_2 \, gdp^2 + \hat{\beta}_3 \, females$$

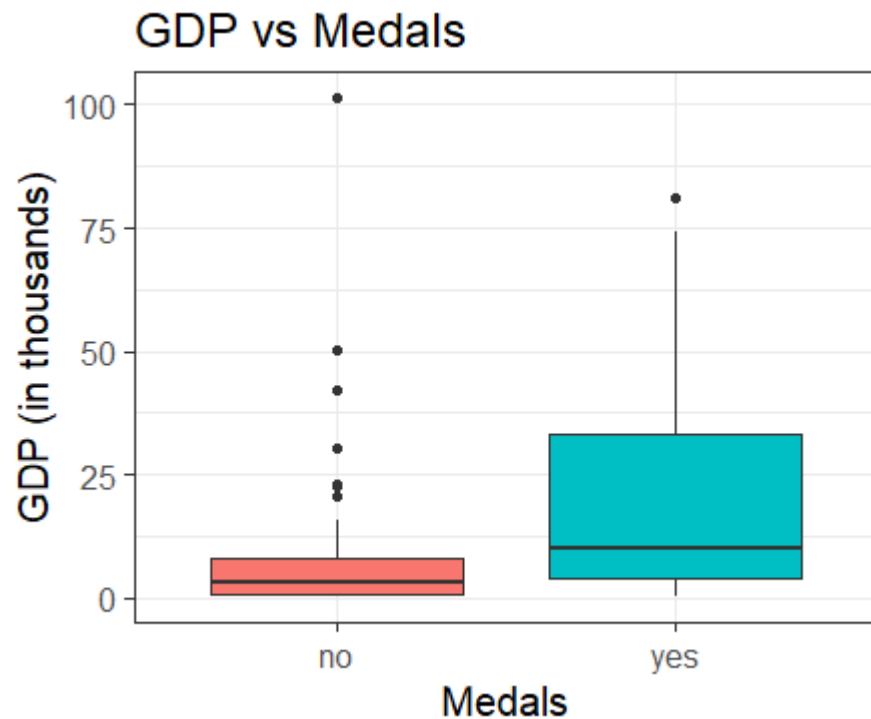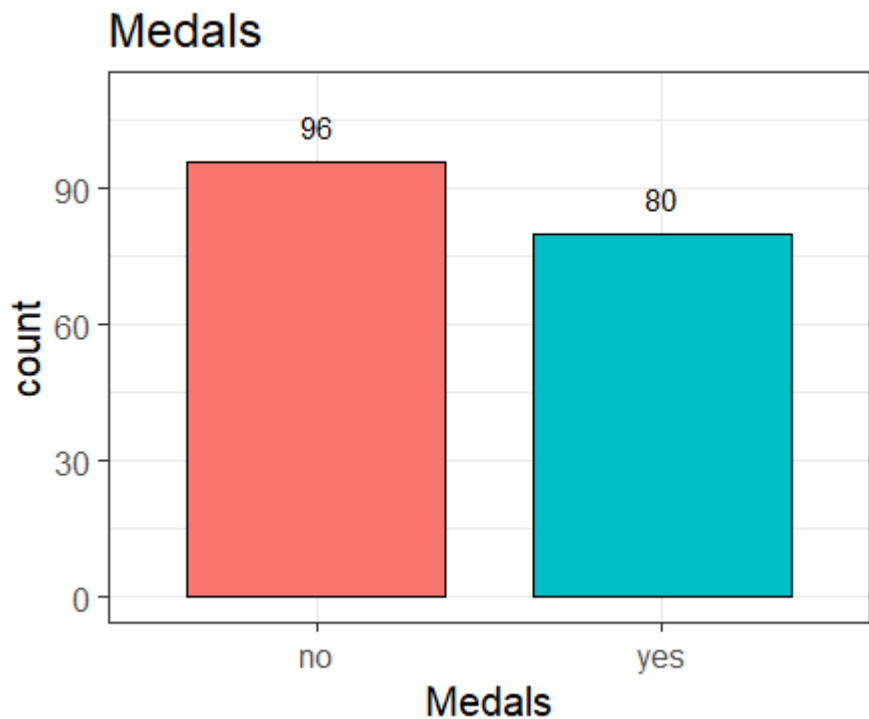|	Coefficient	Std. Error	t value	P-value
(Intercept)	0.99	0.55	1.81	0.072
gdp	15.84	2.48	6.38	1.6e-9
gdp^2	-11.02	2.48	-4.44	1.6e-5
females	0.018	0.012	1.48	0.14

Diagnostic Plots

NEW
Research Question

Which of the following factors are associated with **whether or not a country earned at least 1 medal** in the 2016 Olympics?

- GDP per capita
- Number of athletes per population
- Median age of athletes
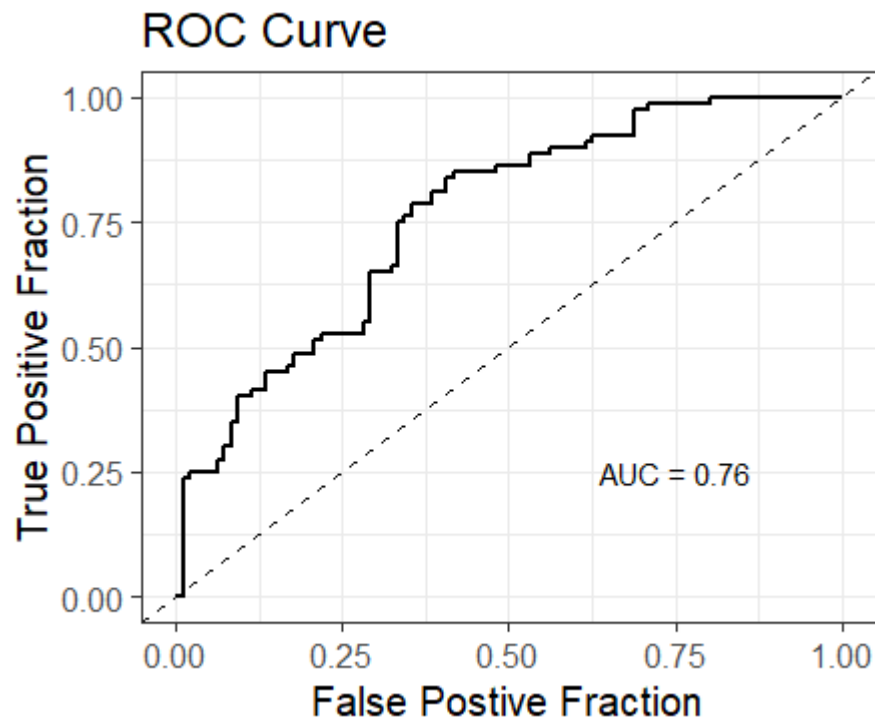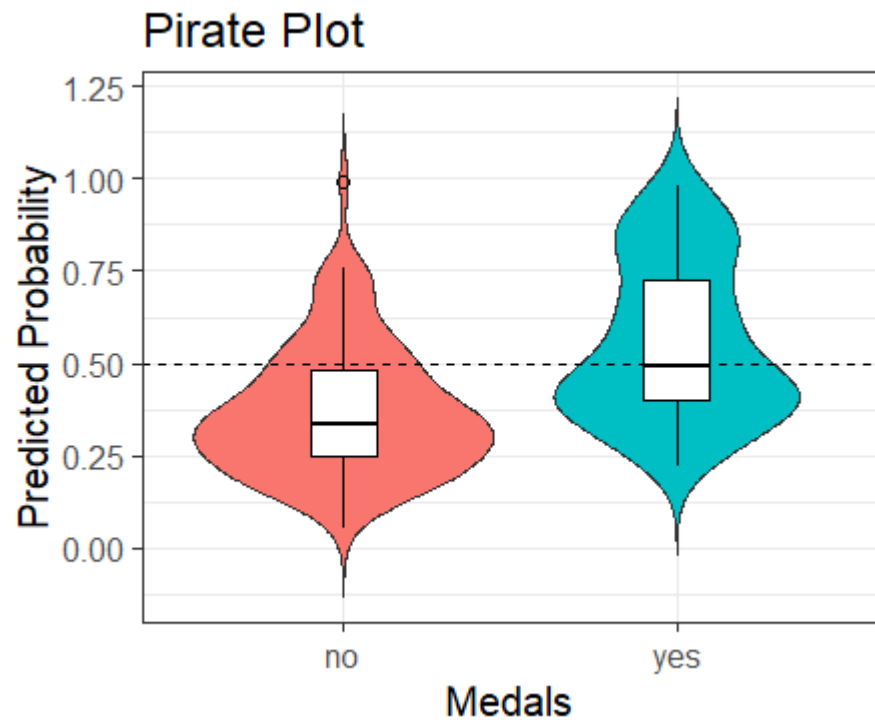- Percentage of female athletes

Exploratory Plots

Logistic Model

$$\hat{p}(X) = \frac{e^{\widehat{\beta}_0 + \widehat{\beta}_1\, gdp + \widehat{\beta}_2\, athletes + \widehat{\beta}_3\, age + \widehat{\beta}_4\, females}}{1 + e^{\widehat{\beta}_0 + \widehat{\beta}_1\, gdp + \widehat{\beta}_2\, athletes + \widehat{\beta}_3\, age + \widehat{\beta}_4\, females}}$$

	Coefficient	Std. Error	z value	P-value
(Intercept)	-3.79	1.75	-2.16	0.031
gdp	0.052	0.014	3.77	1.6e-4
athletes	-0.10	0.073	-1.42	0.16
age	0.16	0.070	2.25	0.024
females	-0.019	0.10	-1.64	0.10

Visualizations

Conclusions

- GDP per capita is associated with the **total number** of medals a country earned
- GDP per capita and the median age is associated with whether a country earned **at least 1** medal

References

- Kaggle: *2016 Olympics in Rio de Janeiro* (https://www.kaggle.com/rio2016/olympic-games)
- Journal of Statistics Education: *A Classroom Investigation of the Effect of Population Size and Income on Success in the London 2012 Olympics* (http://jse.amstat.org/v22n2/carter.pdf)
- An Introduction to Statistical Learning with Applications in R