Winning the Olympics!

Using Data Analysis to predict medal counts

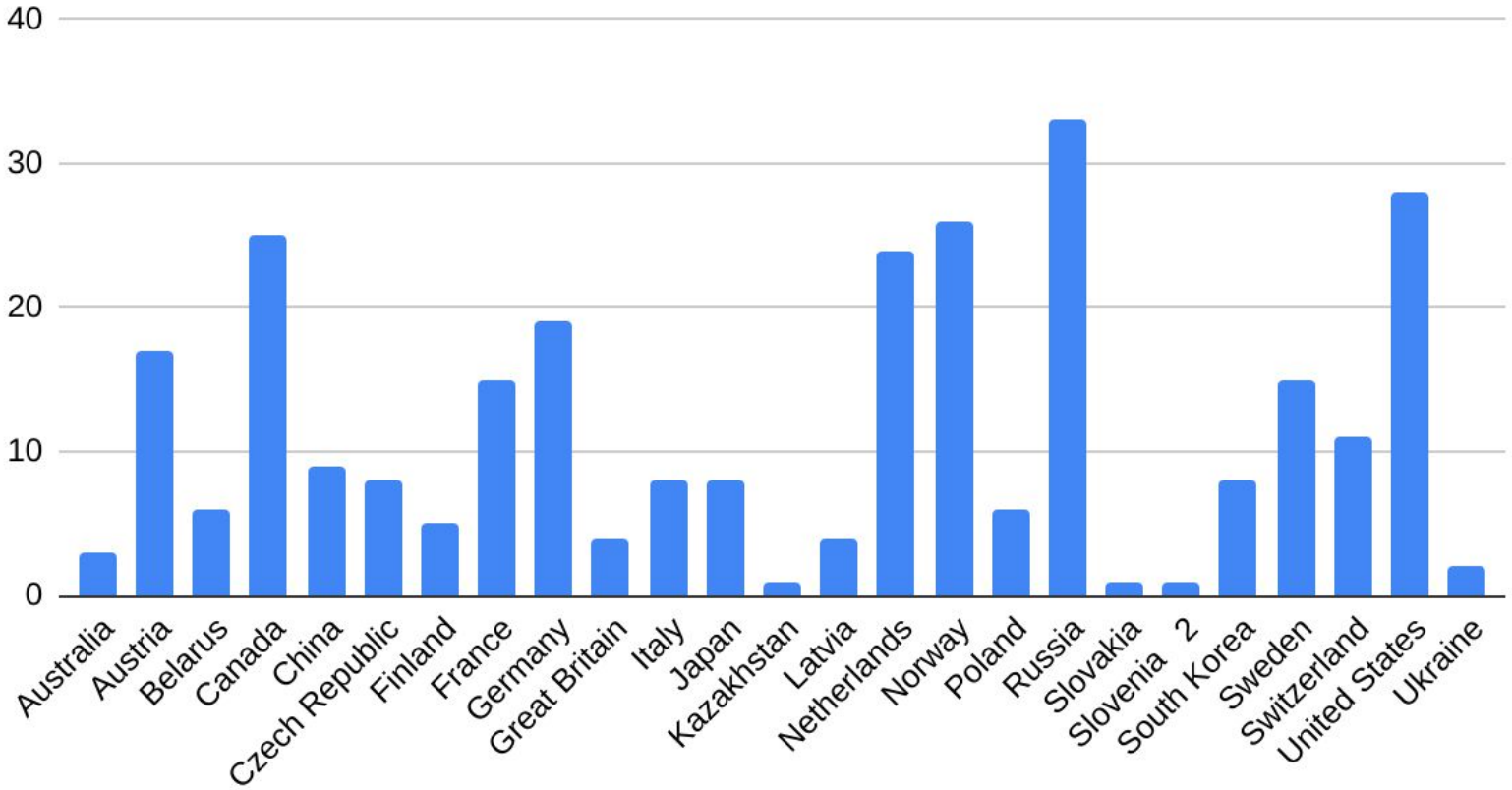
Objective of this project: Use data analysis to find which variable is the best predictor for how a country will do in the Olympics: Previously earned medals? GDP? Latitude?

Use this **data** for the project.

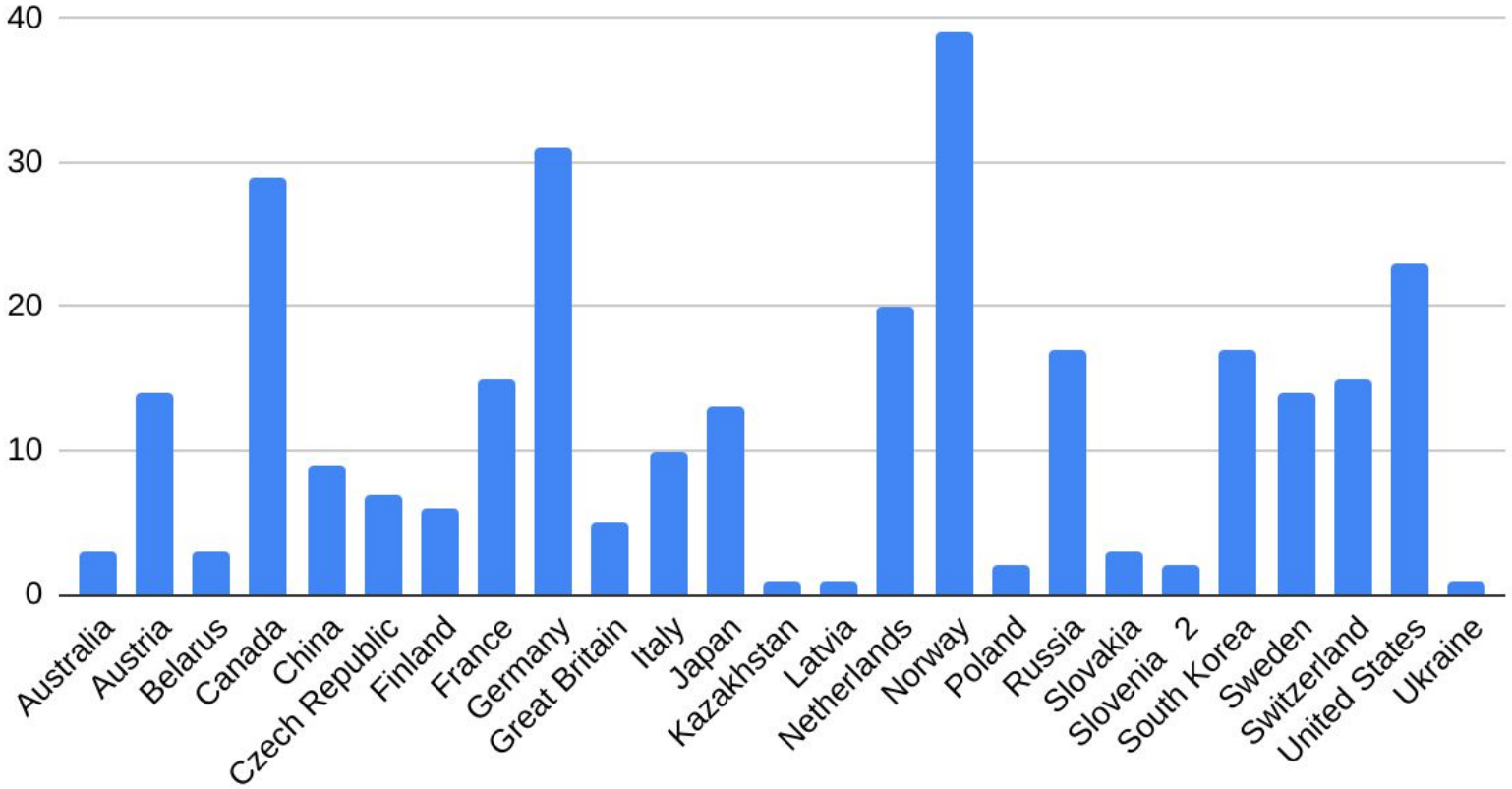
Country	GDP (trillions of USD)	Medals (2018)	Medals (2014)	Latitude	Medals (2018)
Australia	1.323	3	3	35.31	3
Austria	0.417	14	17	48.20	14
Belarus	0.054	3	6	53.90	3
Canada	1.653	29	25	45.42	29
China	12.24	9	9	39.91	9
Czech Republic	0.216	7	8	50.08	7
Finland	0.252	6	5	60.17	6
France	2.583	15	15	48.86	15
Germany	3.677	31	19	52.52	31
Great Britain	2.622	5	4	51.51	5
Italy	1.935	10	8	41.90	10
Japan	4.872	13	8	35.69	13
Kazakhstan	0.159	1	1	51.17	1
Latvia	0.03	1	4	56.95	1
Netherlands	0.826	20	24	52.37	20
Norway	0.399	39	26	59.95	39
Poland	0.525	2	6	52.23	2
Russia	1.578	17	33	55.75	17
Slovakia	0.096	3	1	48.14	3
Slovenia	0.049	2	1	46.06	2
South Korea	1.531	17	8	37.57	17
Sweden	0.538	14	15	59.33	14
Switzerland	0.679	15	11	46.95	15
United States	19.39	23	28	38.90	23
Ukraine	1.122	1	2	48.38	1

Country	GDP (trillions of USD)	Medals (2018)	Medals (2014)	Latitude
Australia	1.323	3	3	35.31
Austria	0.417	14	17	48.2
Belarus	0.054	3	6	53.9
Canada	1.653	29	25	45.42
China	12.24	9	9	39.91
Czech Republic	0.216	7	8	50.08
Finland	0.252	6	5	60.17
France	2.583	15	15	48.86
Germany	3.677	31	19	52.52
Great Britain	2.622	5	4	51.51
Italy	1.935	10	8	41.9
Japan	4.872	13	8	35.69
Kazakhstan	0.159	1	1	51.17
Latvia	0.03	1	4	56.95
Netherlands	0.826	20	24	52.37
Norway	0.399	39	26	59.95
Poland	0.525	2	6	52.23
Russia	1.578	17	33	55.75
Slovakia	0.096	3	1	48.14
Slovenia 2	0.049	2	1	46.06
South Korea	1.531	17	8	37.57
Sweden	0.538	14	15	59.33
Switzerland	0.679	15	11	46.95

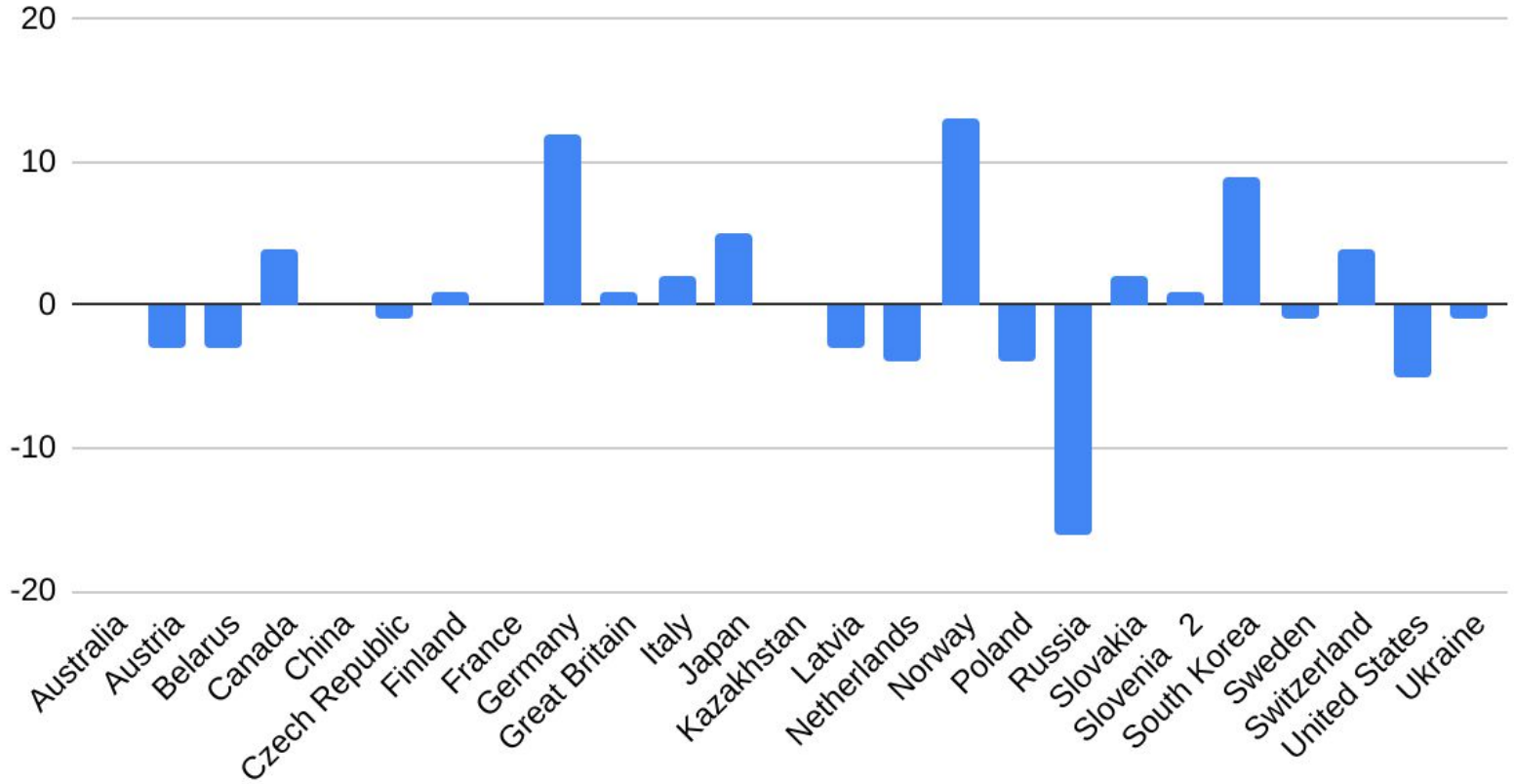
2014 Olympics medal count



2018 Olympics medal count



Olympic medal Count change (2014 - 2018)



2014 Olympics
Sochi, Japan

Yuna Kim
Flikr


DYNASTAR
SOCHI 2014

2018 Olympics
PyeongChang, China

Russia was barred from competing in the 2018 Winter Olympics following a doping scandal dating back to the 2014 Winter Olympics in Sochi. Russia denies the allegations. But some Russian athletes are still being allowed to compete as “Olympic Athletes from Russia,” or “OAR” for short. Feb 15, 2018

https://time.com › Sports › olympics 2018

Why Is There An OAR Team in the 2018 Winter Olympics? | Time

Time.com

2018 Pairs figure skating podium. Wikipedia

CNN.com

Medal Count B & W Plots for 2014 and 2018

- Use ***this*** box and whisker plot tool to model the Medal Count data from each year.
- Upload a screenshot of your box and whisker plot below, 5NumSum should be included in screenshot *(3 points per b&w plot)*

Box and Whisker Plot for 2014 Medals

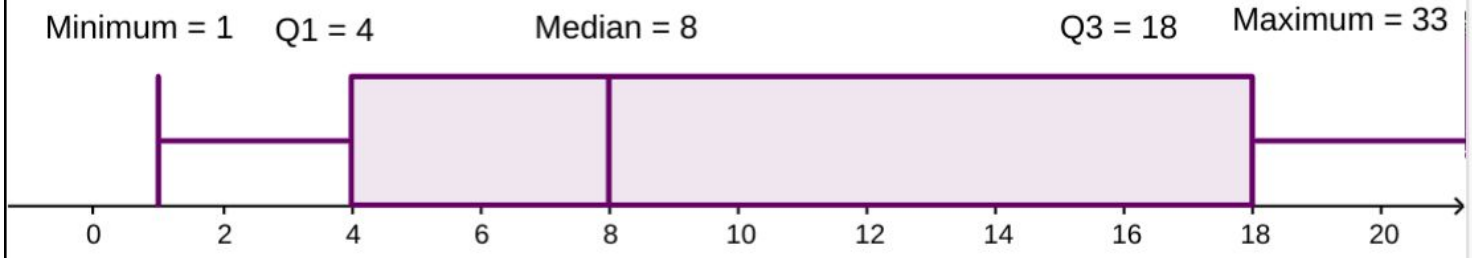
PLEASE SEE NEXT SLIDES
SLIDE 16 and SLIDE 17

Box and Whisker Plot for 2018 Medals

2014 Olympic Medal Count Data

Mean = 11.48

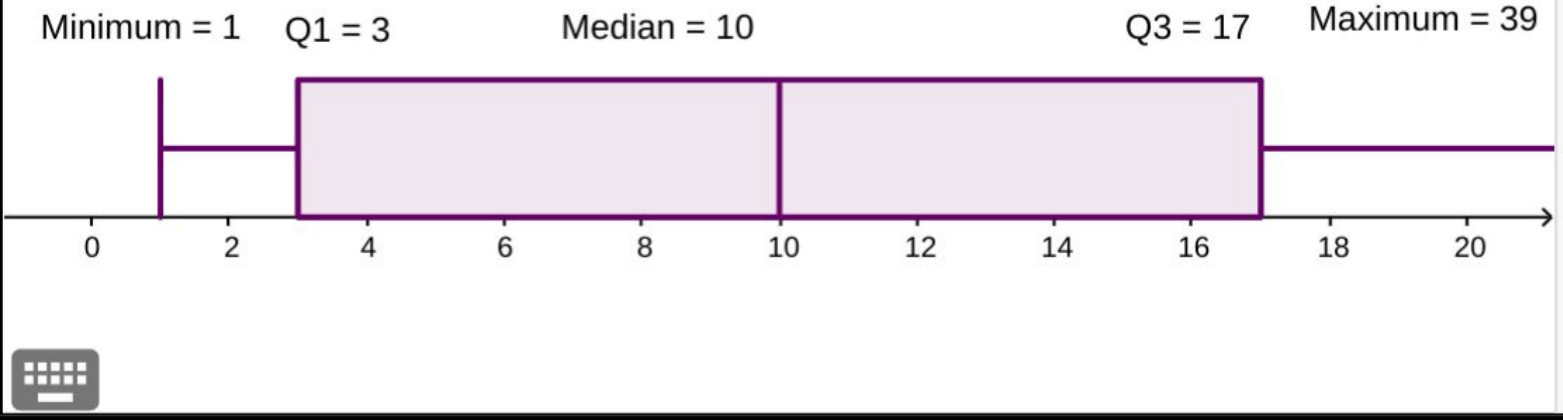
Mode = {8}



	A
16	11
17	15
18	15
19	17
20	19
21	24
22	25
23	26
24	28
25	33

Mean = 12

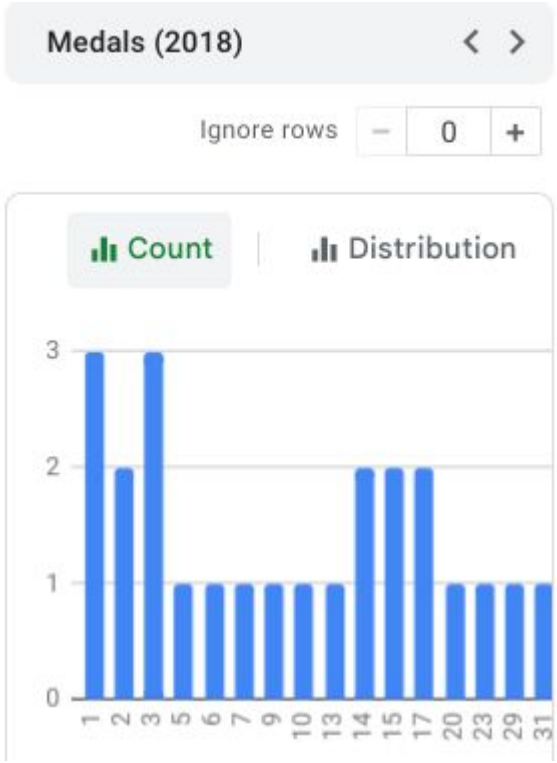
Mode = {1, 3}

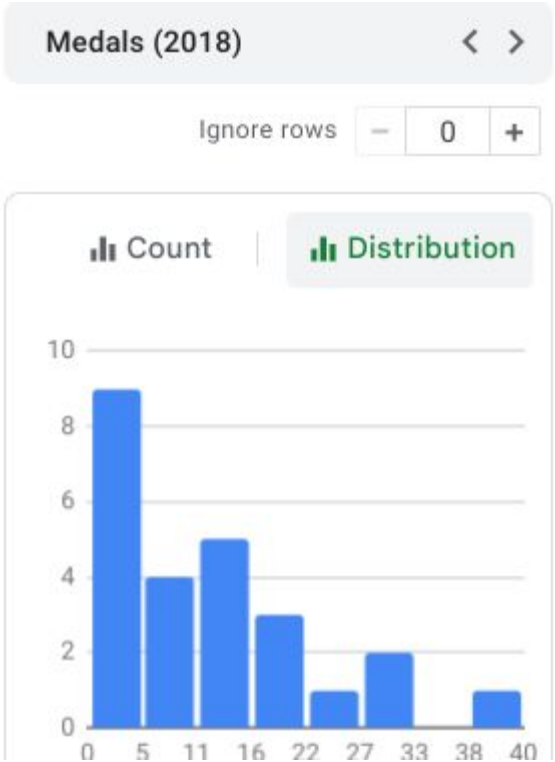


	A
16	14
17	15
18	15
19	17
20	17
21	20
22	23
23	29
24	31
25	39

2018 Olympic Medal Count Data

2018 Medal Statistics





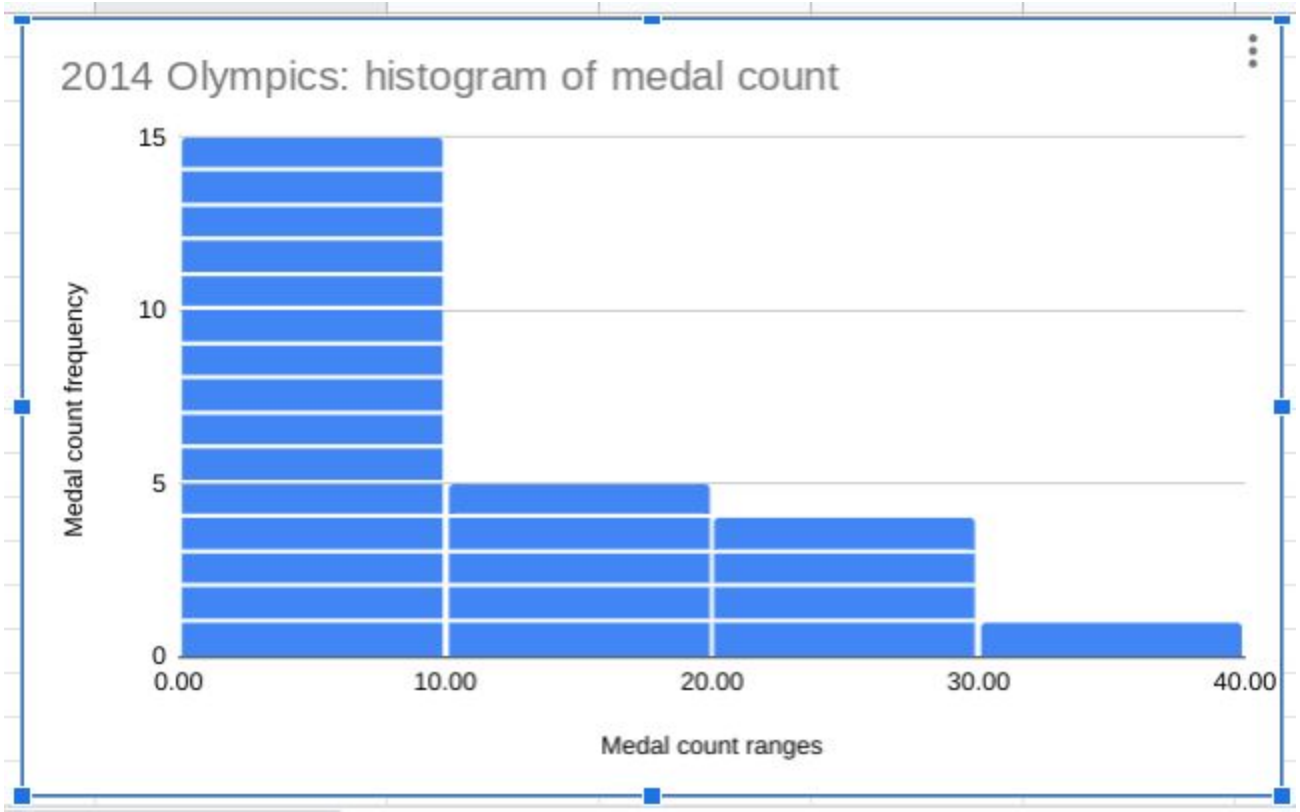
Total rows	1,000
Empty cells	974
Unique values	18
Sum	300
Average	12
Median	10
Minimum value	1
Maximum value	39

Column Statistics in Google Sheets

	A	B
1	1	
2	1	
3	1	
4	2	
5	2	
6	3	Q1 = 3
7	3	
8	3	
9	5	
10	6	
11	7	
12	9	
13	**10**	Median

13	**10**	Median
14	13	
15	14	
16	14	
17	15	
18	15	
19	17	
20	17	Q3 = 17
21	20	
22	23	
23	29	
24	31	
25	39	

IQR = Q3 (17) - Q1 (3) = 10


2014 Olympics: histogram of medal count
15
10
5
0
Medal count frequency
0.00
10.00
20.00
30.00
40.00
Medal count ranges

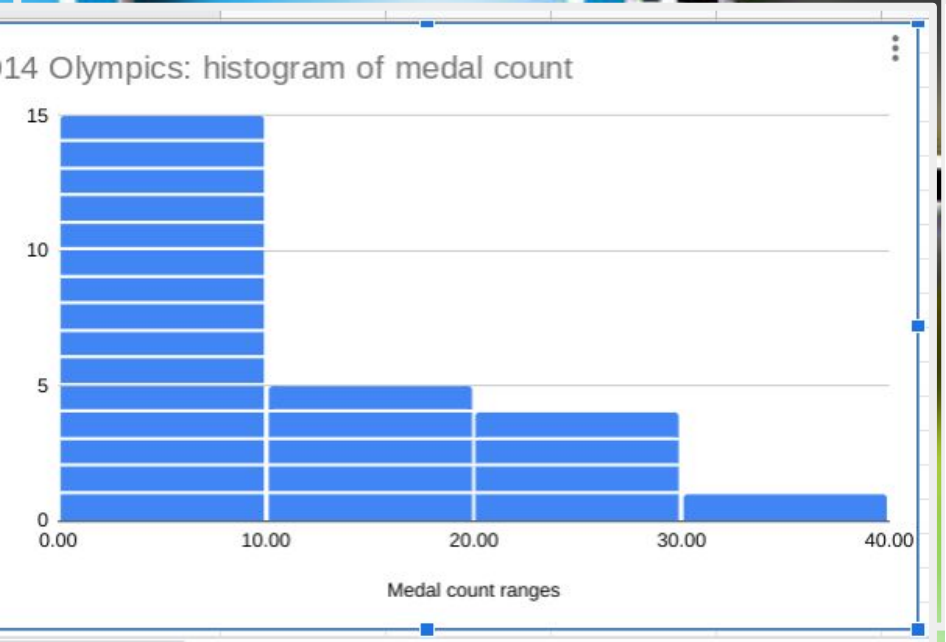
Country	Medals (2014)
Kazakhstan	1
Slovakia	1
Slovenia 2	1
Ukraine	2
Australia	3
Great Britain	4
Latvia	4
Finland	5
Belarus	6
Poland	6
Czech Republic	8
Italy	8
Japan	8
South Korea	8
China	9
Switzerland	11
France	15
Sweden	15
Austria	17
Germany	19
Netherlands	24
Canada	25
Norway	26

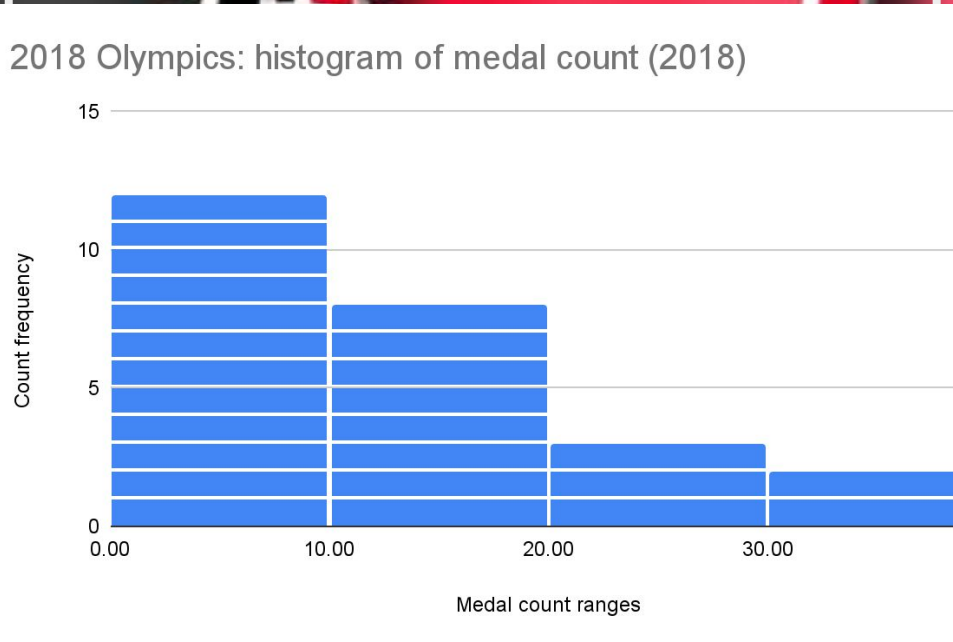
Country	Medals (2018)
Kazakhstan	1
Latvia	1
Ukraine	1
Poland	2
Slovenia 2	2
Australia	3
Belarus	3
Slovakia	3
Great Britain	5
Finland	6
Czech Republic	7
China	9
Italy	10
Japan	13
Austria	14
Sweden	14
France	15
Switzerland	15
Russia	17
South Korea	17
Netherlands	20
United States	23
Canada	29
Germany	31

Histograms of Medal Counts

Use **google sheets** to create a histogram for 2014 and 2018.

- ➔ Upload a screenshot of each histogram.
- ➔ Use a minimum value of 0 and a 'bucket size' of 10. *(1 pt each)*
- ➔ Correctly label your axes as well. *(1 pt each)*





Analyzing your Data

In which year did more countries reach a higher medal count? *(1 pt)* ***2018***	According to the histogram, 2 countries won the top bracket of 30 to 40 medals in 2018. Only 1 country (Russia) did so in 2014. They were banned for doping in 2018. **According to slide 6 (change in medal count 2018-2014) 11 countries had more medals in 2018; 10 won fewer. Germany, Norway and South Korea did significantly better in 2014 (slide 6). Russia had much fewer in 2018. The median count of 10 was higher in 2018 than in 2014, when the median medal count was 8.**
How do you know? ***Explain***.(2 pts)	In 2014 Japan, the total medal count was 287. In China in 2018, total 300 medals were awarded. More countries won at least one and up to 10 medals in Japan 2014 according to the histogram. We know more countries won medals because the median medal count was higher in 2018.
In which year did more countries receive less than 10 medals? (1 pt)	More countries received less than 10 medals in but at least one medal in 2014. There were **15 countries in this range in 2014**, and 12 in 2018.
How do you know? ***Explain***. (2 pts)	Data table 19 shows that each country won at least one medal, so the histogram gives the correct impression that was more opportunity on the low end of the medal county in 2014.
IQR for 2014, *show work*. (2 pts)	IQR = Q3 (18) - Q1 (4) = 14 (slide 16 box and whisker plot)

Kazakhstan	1	
Latvia	1	
Ukraine	1	
Poland	2	
Slovenia 2	2	
Australia	3	**Q1 3**
Belarus	3	
Slovakia	3	
Great Britain	5	
Finland	6	
Czech Republic	7	
China	9	
Italy	**10**	**Median 10**
Japan	13	
Austria	14	
Sweden	14	
France	15	
Switzerland	15	**Q3**
Russia	17	**17**
South Korea	17	
Netherlands	20	

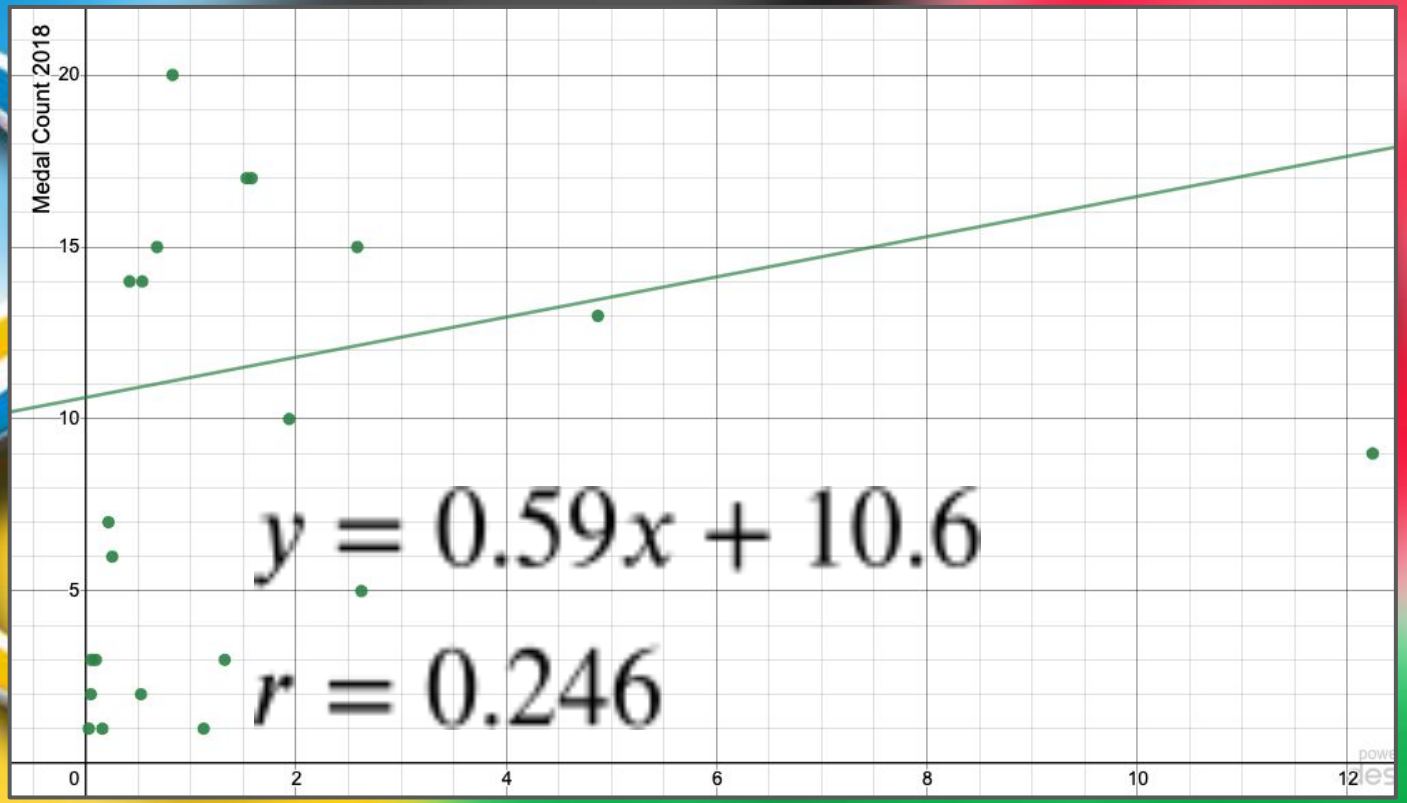
2018 data

IQR = Q3(17) - Q1(3) = 14

GDP vs. 2018 Medal Counts

This is a scatter plot and line of best fit comparing the **GDP** (***how wealthy a country is***) to how many medals the country earned in the 2018 Olympics.

The equation of line of best fit and correlation coefficient are given.



Use the previous graph, equation, and correlation coefficient to answer these questions.

What does the data tell you about the relationship between **GDP** and performance in the 2018 Winter Olympics? ***Explain in full sentences***. *(3 pts)* The best fit formula shows a positive r value of 0.246. This is considered a weak positive correlation. GDP is weakly associated with the number of olympic medals won in 2018.

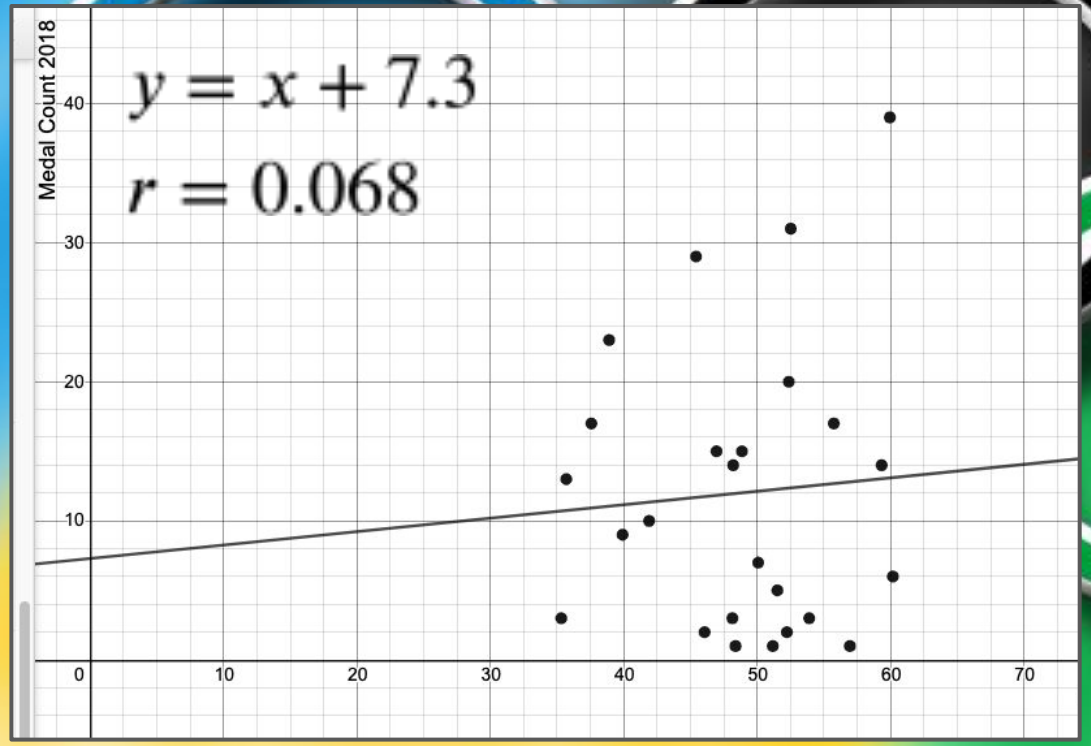
What do the x and y represent in this scatter plot? ***Explain in full sentences***. *(2 pts)* The x-coordinate represents the GDP in trillions of dollars. The y axis records the medal count in 2018.

What does the ***slope*** of 0.59 represent? ***Explain*** **in a full sentence.** *(2 pts) The slope represents the medal count expected per trillion dollars of GDP.*

The GDP of Hungary was approximately $0.158 trillion in 2018. Based on this, how many medals would you predict they won in the 2018 Winter Olympics using your line of best fit? ***Show all work*** → (2 pts)	Using the best fit equation y = 0.59x + 10.6, the medal prediction would be (0.59)(0.158) + 10.6 = 0.0932 + 10.6 = 10.6932. This is approximately 11 medals.
How confident are you in this prediction? Explain why in full sentences. *(Saying you 'checked your work' isn't sufficient)* (2 pts)	The best fit equation predicts 11 medals. There is low confidence because the correlation factor is low.
Look online here: how many medals did Hungary **actually** win? (1 pt)	Hungary won 1 medal.
Was your prediction right or wrong? Explain in full sentences. (2 pts)	The actual medal count was quite lower than prediction of 11 medals. The prediction of low confidence was appropriate.
Are you surprised or not? Explain in full sentences. (2 pts)	I am not surprised because few countries can earn more than 10 medals in the olympics. The data was loosely correlated, so this met the expectation that the estimate would not be accurate.

What about other factors?

Here is the scatter plot and line of best fit comparing the ***medal count of 2018*** with the ***latitude*** of the country.



Why might latitude affect the medal count of the Winter Olympics? ***Explain.***

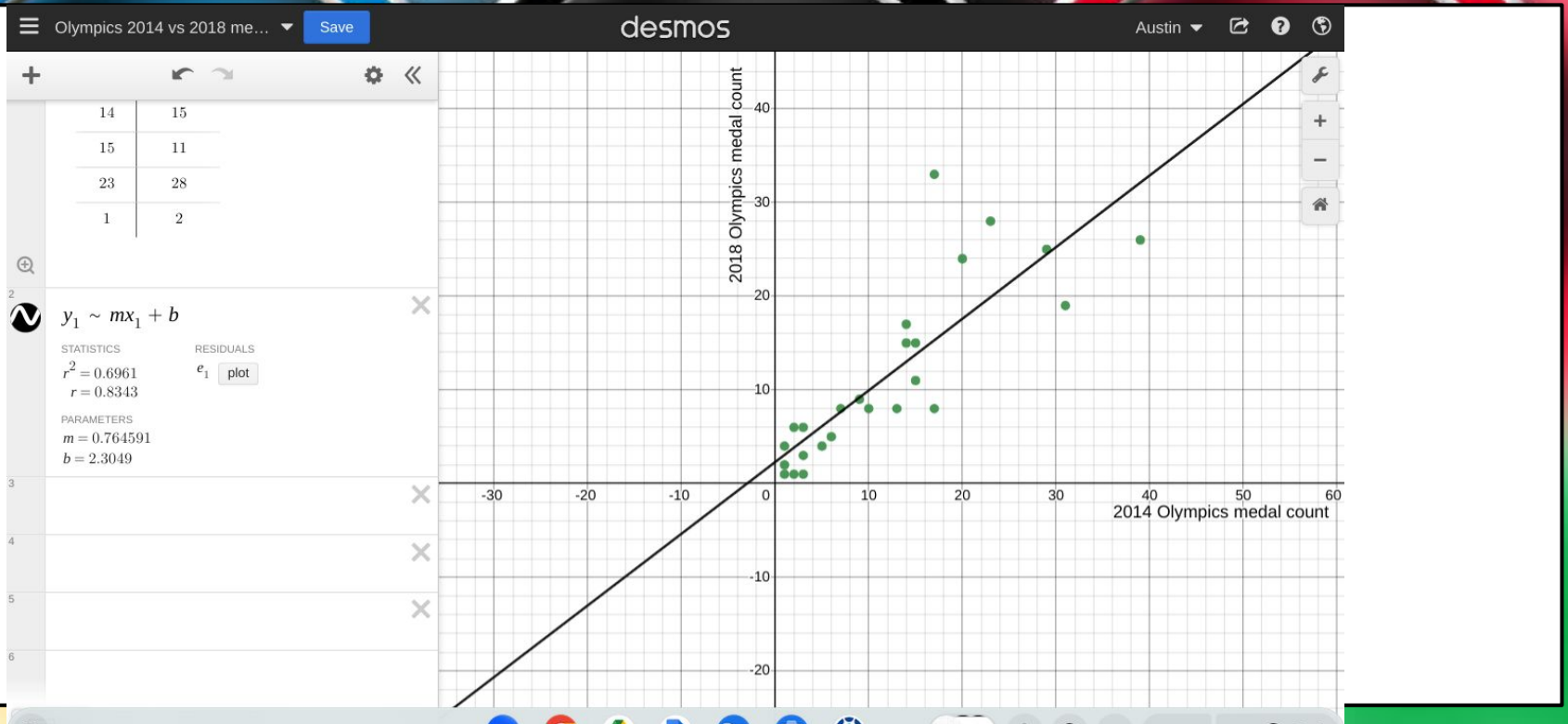
Higher latitude, more northerly countries have more winter weather and more expertise in winter sports, so they might be expected to win more medals.

How strong of an influence does latitude seem to have? ***Explain*** how you know.

The best fit data says this is not a reliable indicator. There is a very weak correlation

Create your own...

- Use the Data Sheet on slide 2 and **Desmos** to create your own scatter plot, line of best fit, and correlation coefficient. You will use the scatter plot and Desmos info on the next slide.
- **Use Medals won in 2014 as your x_1 values *vs.* Medals won in 2018 as your y_1 values.**



Use your scatter plot and line of best fit to answer the questions below:

What is the ***equation*** of your line of best fit comparing medals won in 2014 vs 2018 Winter Olympics? *(2 pts)*

$y = mx + b$

M = 0.764591, b=2.3049

What is your **correlation coefficient** for the comparison of 2014 vs 2018 medals? *(1 pt)*

$r = 0.8343$

Based on this exploration, which variable - **GDP**, **latitude**, or **medals earned in 2014** - does the best job of predicting performance in the Olympics? **Explain**, using full sentences *(3 pts)*

Prior performance in winning medals was by far the best indicator of the future medal count. This best fit equation predicts a country is expected to earn 2 medals plus 76% of prior earnings. The correlation coefficient of r = 0.83 is considered to be a strong connection indicator. The r value of GDP was 0.246 and that of latitude was 0.068. These are much weaker prediction models.

Honors Only Slide

For each of the three linear models, find an example of a country whose performance at the Olympics was close to the model's predictions:

GDP vs. Medal count 2018	
Latitude vs. Medal count 2018	
Medal count 2014 vs. 2018	

For each of the three linear models, find an example of a country whose performance at the Olympics was ***not*** close to the model's predictions:

GDP vs. Medal count 2018	
Latitude vs. Medal count 2018	
Medal count 2014 vs. 2018	