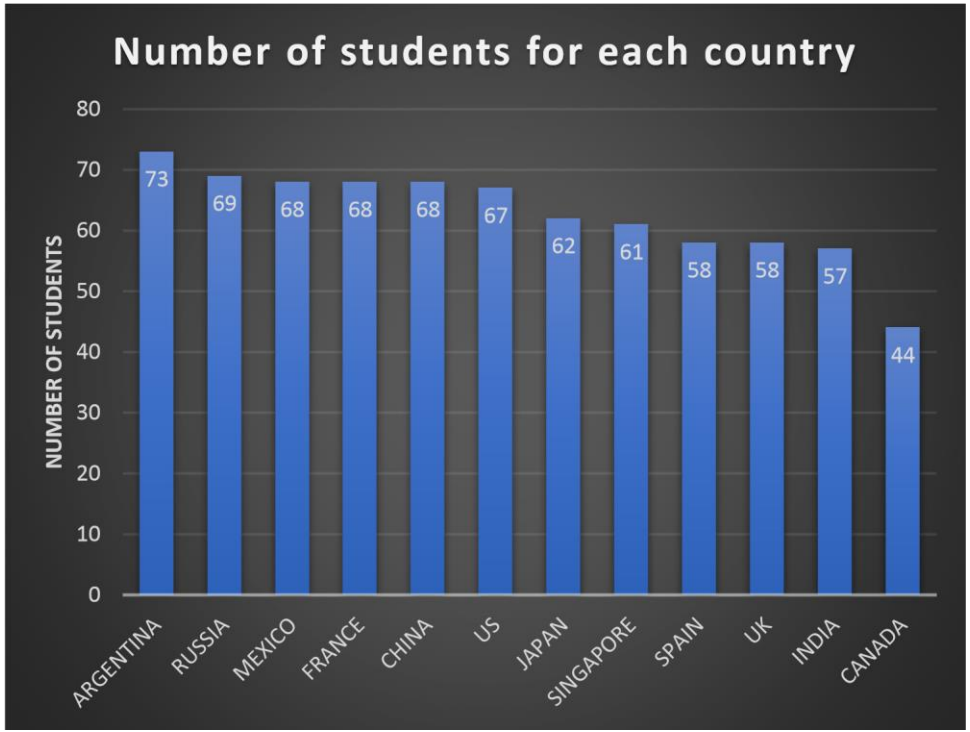
1.What are the most common countries where students live?



In this case try to find which country has the maximum number of students registered in Udacity database, so First, count numbers of students for each country using PivotTable. Then, Find the maximum number in the

PivotTable using this formula: =MAX(GETPIVOTDATA("Count of ARGENTINA",A3),GETPIVOTDATA("Count of SPAIN",A3),GETPIVOTDATA("Count of US",A3),GETPIVOTDATA("Count of UK",A3),GETPIVOTDATA("Count of RUSSIA",A3),GETPIVOTDATA("Count of JAPAN",A3),GETPIVOTDATA("Count of SINGAPORE",A3),GETPIVOTDATA("Count of MEXICO",A3),GETPIVOTDATA("Count of FRANCE",A3),GETPIVOTDATA("Count of INDIA",A3),GETPIVOTDATA("Count of CANADA",A3),GETPIVOTDATA("Count of CHINA",A3))

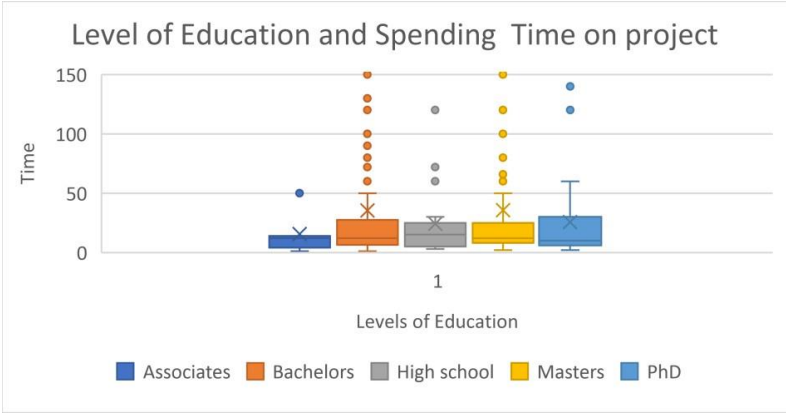
So the max number of students are belong to Argentina

2.Does the level of education have effect on number of hours spending on projects?

Comparing using 5 number summaries find Bachelors student has the maximum value in spending time working on project this value is extremely high can consider as outlier value.

Depends on outlier values that effect the Mean of data which is sensitive measure, the best measure to compare these 5 groups is median(Q2) which is resistant measure. So, High school student are spending more time on project than any other group may they need more help for working on project rather than other levels' students are more independent.

Number of hours spend/ project	Associates	Bachelors	High school	Masters	PhD
Minmum	50	1000	120	800	300
Q1	4.5	7	7.5	8	6
Q2	9	12	15.5	12	10
Q3	12	25	24	25	30
Maximum	1	1	3	2	2



Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Associates	11	171	15.54545	310.8727
Bachelors	265	9390	35.43396	7209.58
High school	19	462	24.31579	875.4503
Masters	296	10518	35.53378	6922.595
PhD	69	1761	25.52174	1842.312

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	11700.34	4	2925.085	0.468484	0.758898	2.385533
Within Groups	4089639	655	6243.723			
Total	4101339	659				

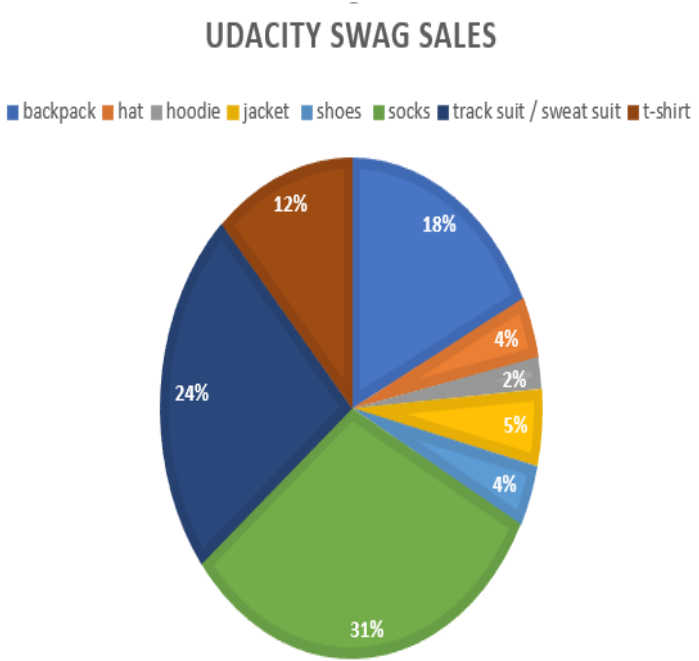
Test hypothesis if there is a significant difference between levels of education in spending time on project or not. For that we need to calculate SD then use ANOVA.

Here P-value > 0.05 and F<1 that mean accept null hypothesis “No significant difference between groups” and reject alternative hypothesis “there is a significant difference between groups”.

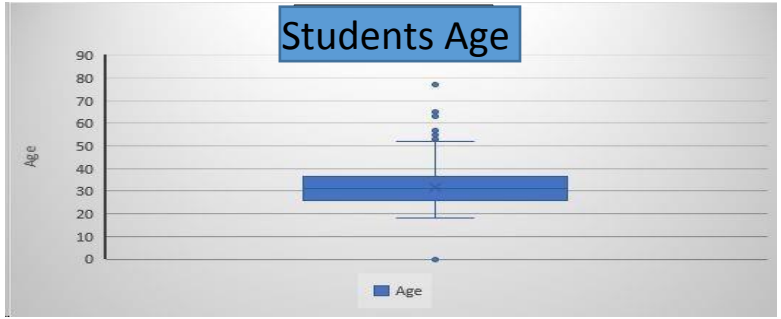
3.what are the best sales of Udacity swag?

Students are buy socks more than any other product. The graph show the percentage of sales for each product.

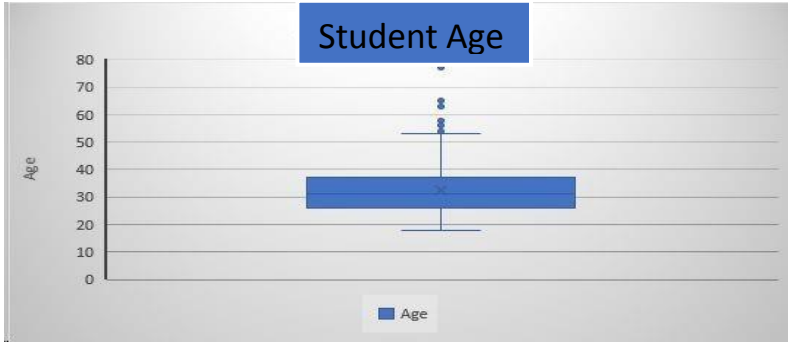
backpack	hat	hoodie	jacket	shoes	socks	track suit /	t-shirt
18%	4%	2%	5%	4%	32%	24%	12%



4.What is the average age of students?



Before cleaning dataset



After cleaning dataset

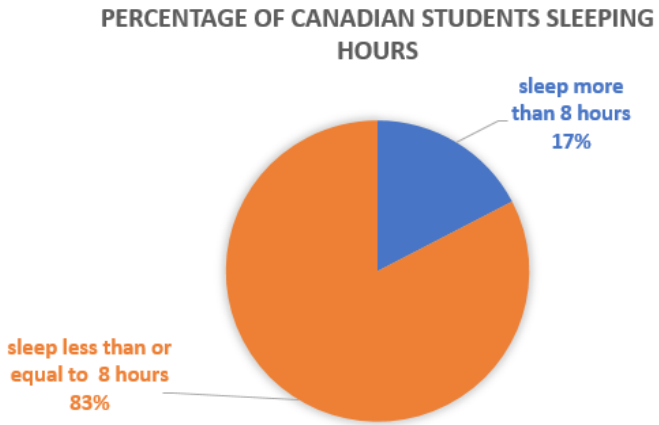
This case help udacity to knew more about its audience, So seek in details about students' age find 7 values related to ages, when start analysis the data there is some outlier "age=0" which is not logical. **Before cleaning dataset**:
Minimum age=0, Average age=31, Maximum age=77
Range=77, Q1=26, Q3=36, IQR=10

After cleaning dataset:
Minimum age=18, Average age=32, Maximum age=77
Range=59, Q1=26, Q3=37, IQR=11

After looking at these values may support decisions related to students' age.

5.Wht is the percentage of Canadian students are sleeping more than 8 hours per night?

mean	7
Median	7
mode	7
SD	1.0530848
Variance	1.1089876
Z-Score	0.94959



Data for sleep hours of Canadian students are standard normally distributed. So to get percentage of students are sleep more than 8 hours calculate

$z\text{-score} = (X - \mu) / \sigma$

The percentage= 17.36% of Canadian students are sleep more than 8 hours

References

N/A