Math 200
Shape Center Spread

Name

Dante Rueda

In this activity, we will use the distance from school dataset. Use this dataset to answer the following questions.

Do most students drive to school? (Your thoughts): I believe most students do to school

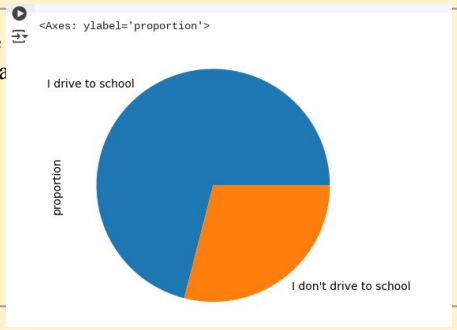
Frequency Table

```
distance_df['travel_method'].val
```

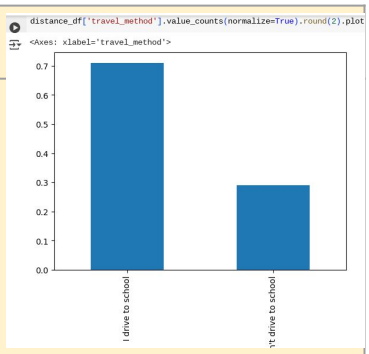
travel_method	proportion
I drive to school	0.71
I don't drive to school	0.29

dtype: float64

Pie Cha



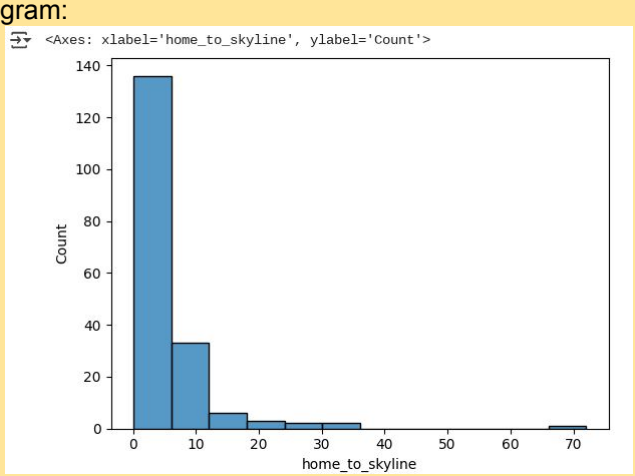
Bar Charts



Answer the question:Most Students travel method to school is by driving.

What do you think is the average distance from students home to skyline? I believe the average distance from students homes to skyline is about 5 miles.

Histogram:

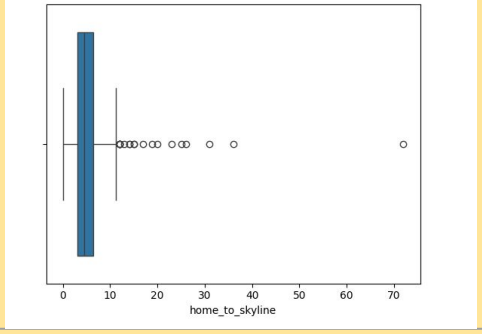



```
#Average distance to school
distance_df['home_to_skyline'].describe().round(2)
```

	home_to_skyline
count	183.00
mean	6.15
std	7.10
min	0.10
25%	3.00
50%	4.50
75%	6.35
max	72.00

Boxplot:

```
sns.boxplot(x = 'home_to_skyline' , data = distance_df)
```

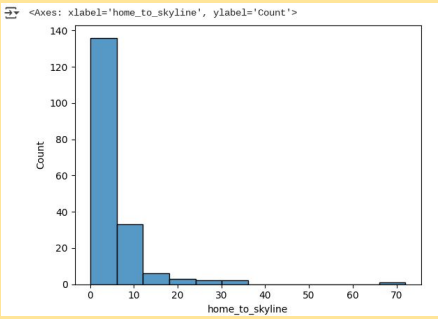


Answer the question:

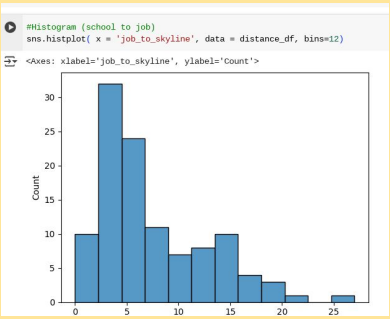
Describe spread of the data.

Compare 3 histograms (distance to school, distance from school to job, job to home). Comment on the shape and the standard deviation of each.

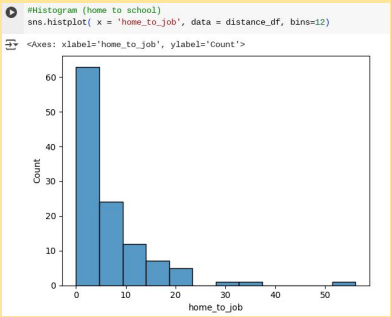
Distance from home to school



Distance from school to job



Distance from job to home



Comment on the shape and the standard deviation of each.

The largest area of distance between each other is the job to skyline, apart from the 72 mile distance from someone's home to skyline.

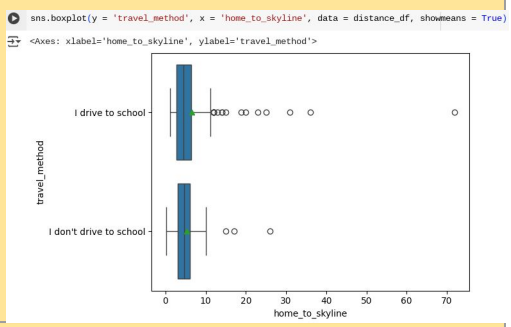
Part 2: Do students who drive to school travel farther on average?

Table:

```
distance_df.groupby('travel_method')['home_to_skyline'].mean()
```

travel_method	home_to_skyline
I don't drive to school	5.323208
I drive to school	6.491473

Side by Side Box Plots



Answer: According to the data, students who drive to school tend to travel more to get to school.

Who do you think has a higher GPA, students who are work part time, full time, or unemployed? I believe either part time or unemployed students have the higher gpas.

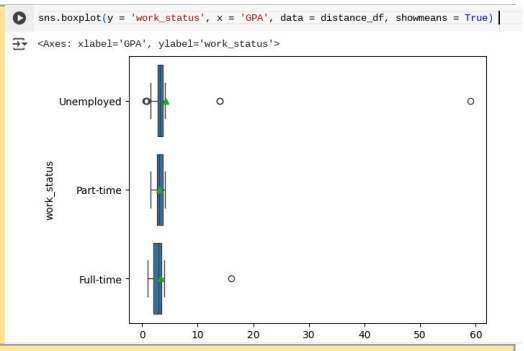
Table:

```
gpa = distance_df.groupby('work_status')['GPA'].mean()
gpa
```

work_status	GPA
Full-time	3.347826
Part-time	3.137432
Unemployed	4.202535

dtype: float64

Side by Side Boxplots



Answer: According to the graph students that are unemployed have better gpas than employed students.

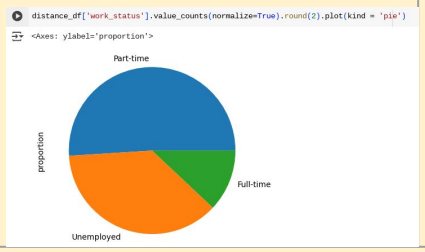
What is most students' employment status? (Your thoughts): I believe that most students employment status' are unemployed.

Frequency Tables

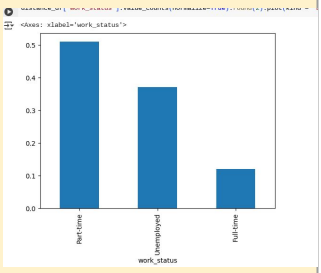
```
distance_df['work_status'].value_counts(normalize=True).round(2)
```

work_status	proportion
Part-time	0.51
Unemployed	0.37
Full-time	0.12

Pie Chart



Bar Charts



Answer the question:Most of the students are part time rather than unemployed

3. Is the there a difference in work status and GPA? Who do you think has the lowest GPA and why? I believe the gpa of part time students is the lowest because there's more part time students to account for than employed and unemployed students

Table:

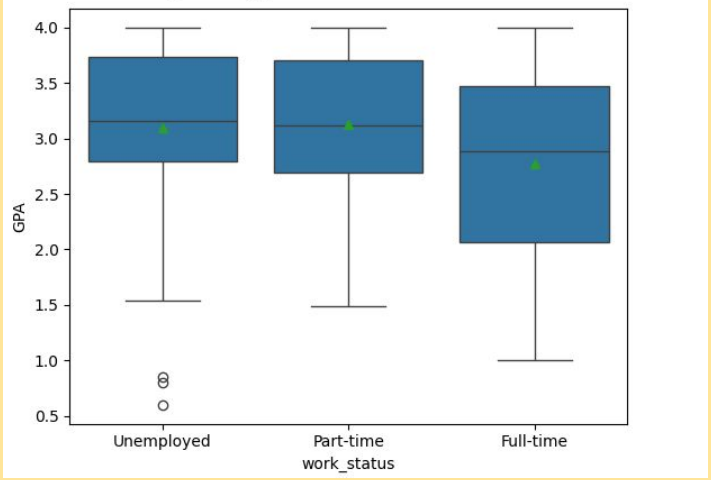
```
distance_df['work_status'].describe().round(2)
```

	work_status
count	183
unique	3
top	Part-time
freq	94

dtype: object

Side by Side Boxplots

```
sns.boxplot(y = 'GPA', x = 'work_status', data = distance_df, showmeans = True)
```



Answer:
With the top being part time the gpa of them being the lowest it shows part time has the worst gpa

Who do you think drives farther to school, male or female students?

Table:	Side by Side Boxplots
Answer:	

Extra Credit:

Using the plotly map, which area do most students located? Create a map