



Seeing Your Data Relationships



Brock Tubre

INSTRUCTOR

Relationships

Visualizing relationships in your data can provide a good general overview, show distribution, and correlation between attributes. Visualizing relationships can also help find outliers and extreme values.

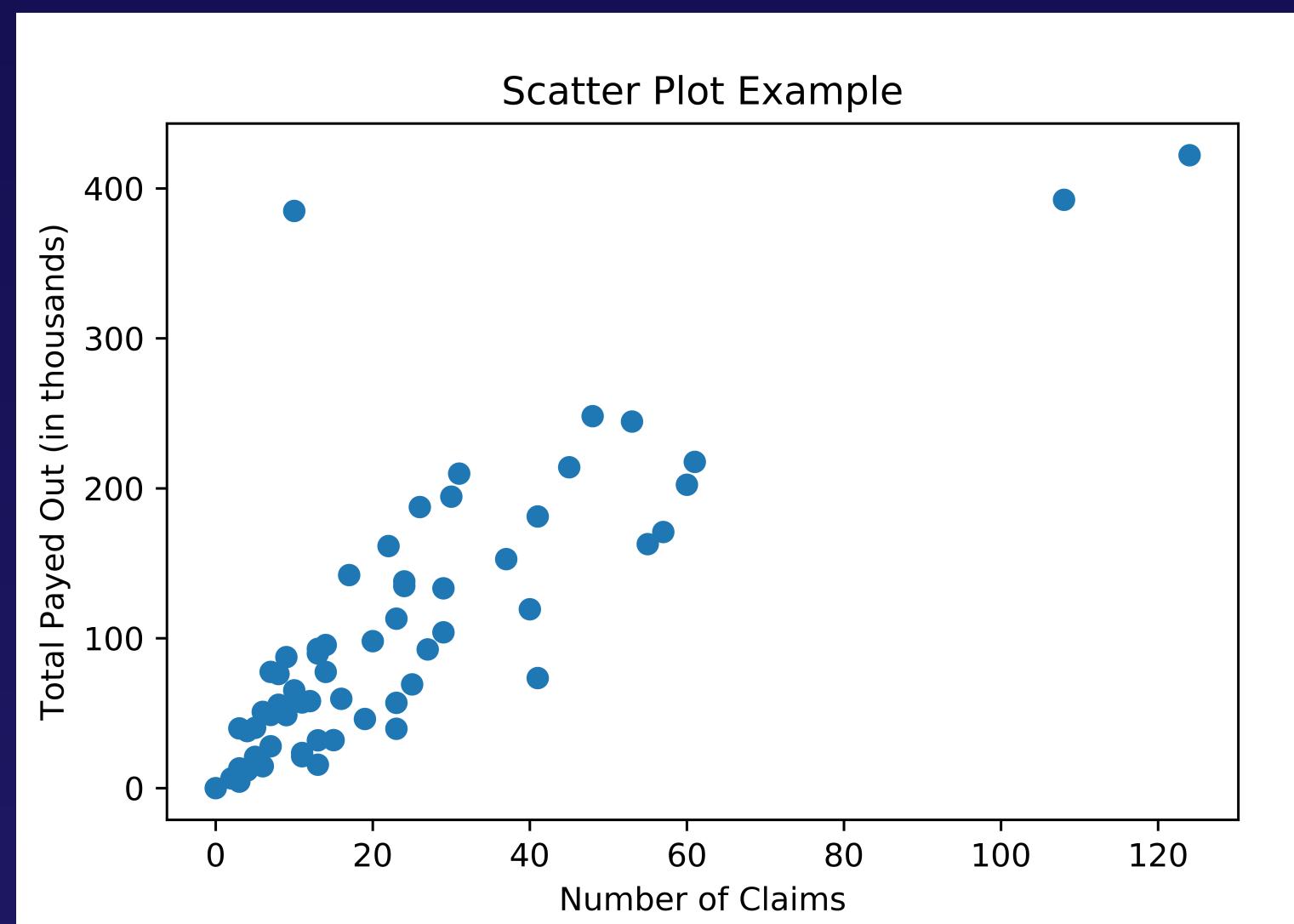


Seeing Your Data - Relationships

1

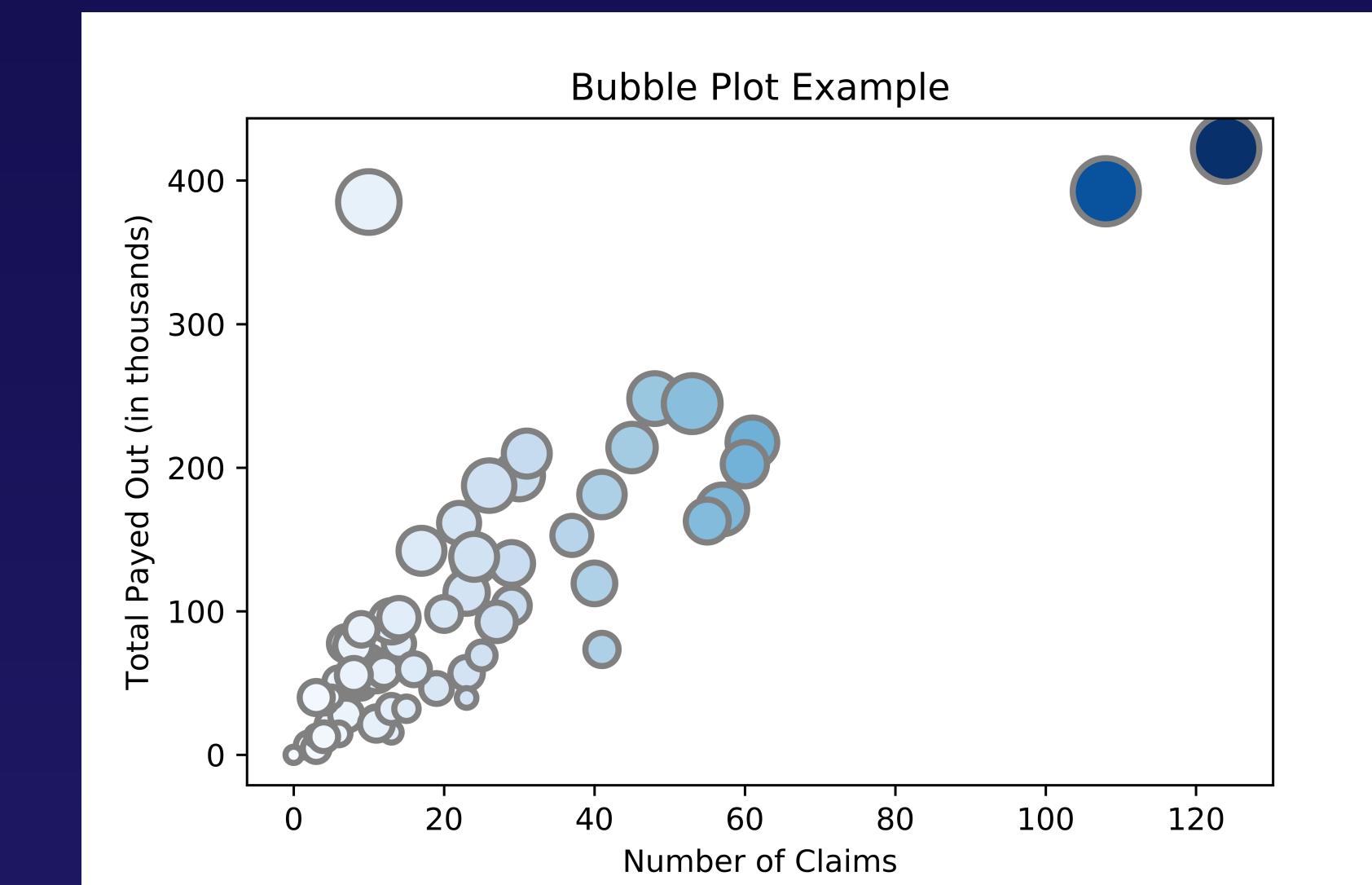
Scatter Plots

Also known as scatter charts. These graphs plot points along the x and y axis for two values.

**2**

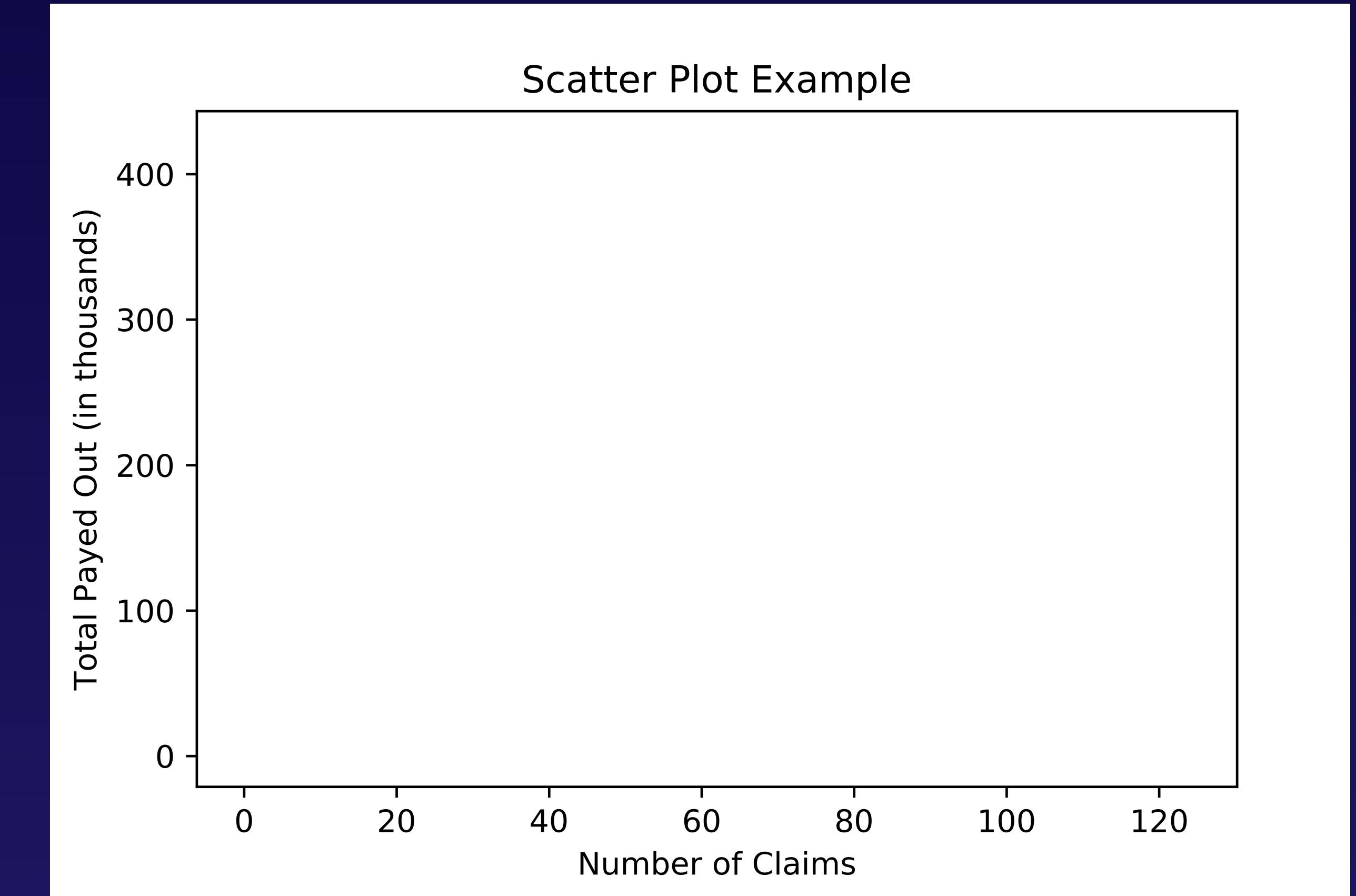
Bubble Plots

Also known as bubble charts. These graphs plot points along the x and y axis for three values. Bubble size is the third value measured.



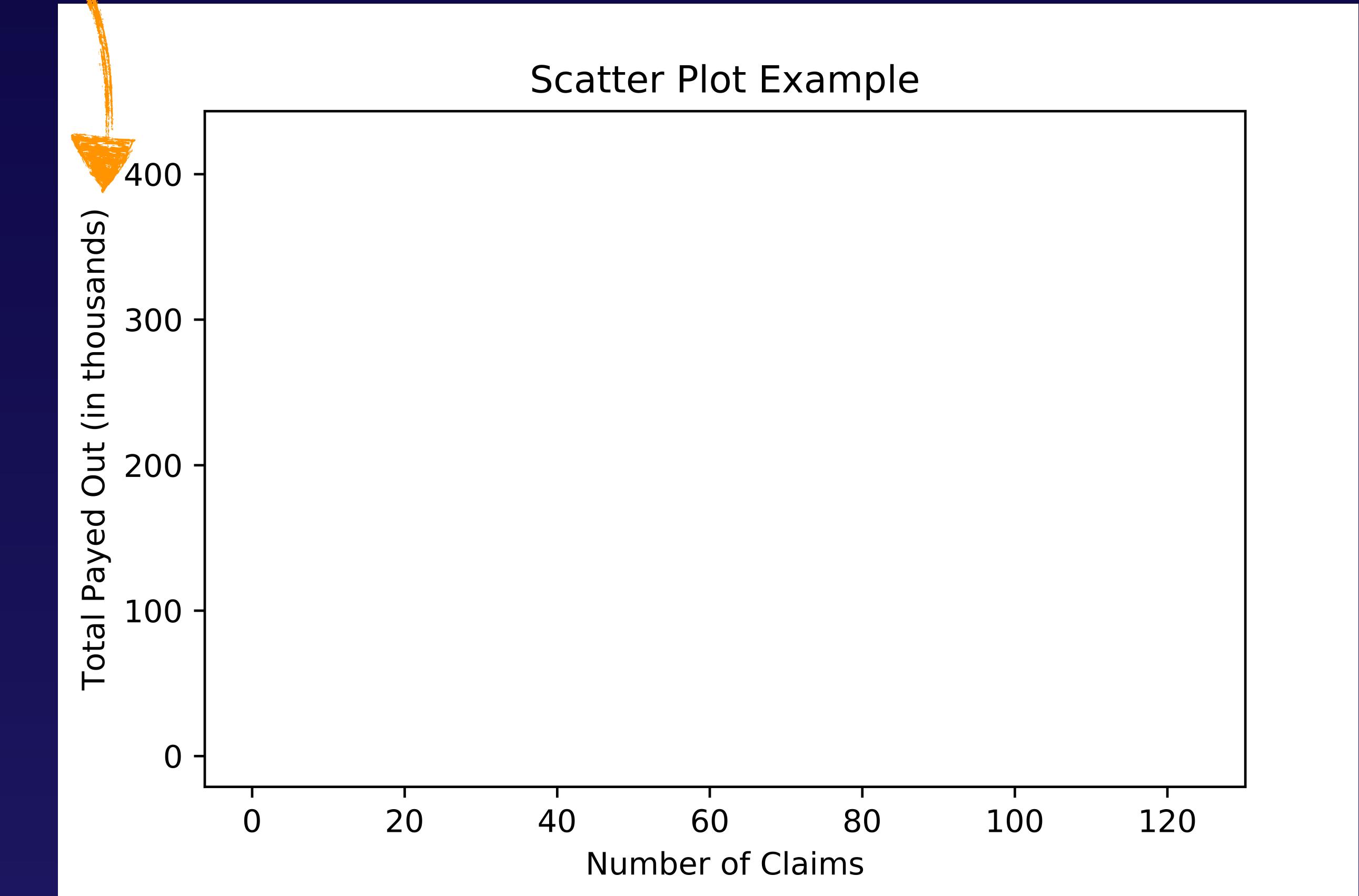
Scatter Plots

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



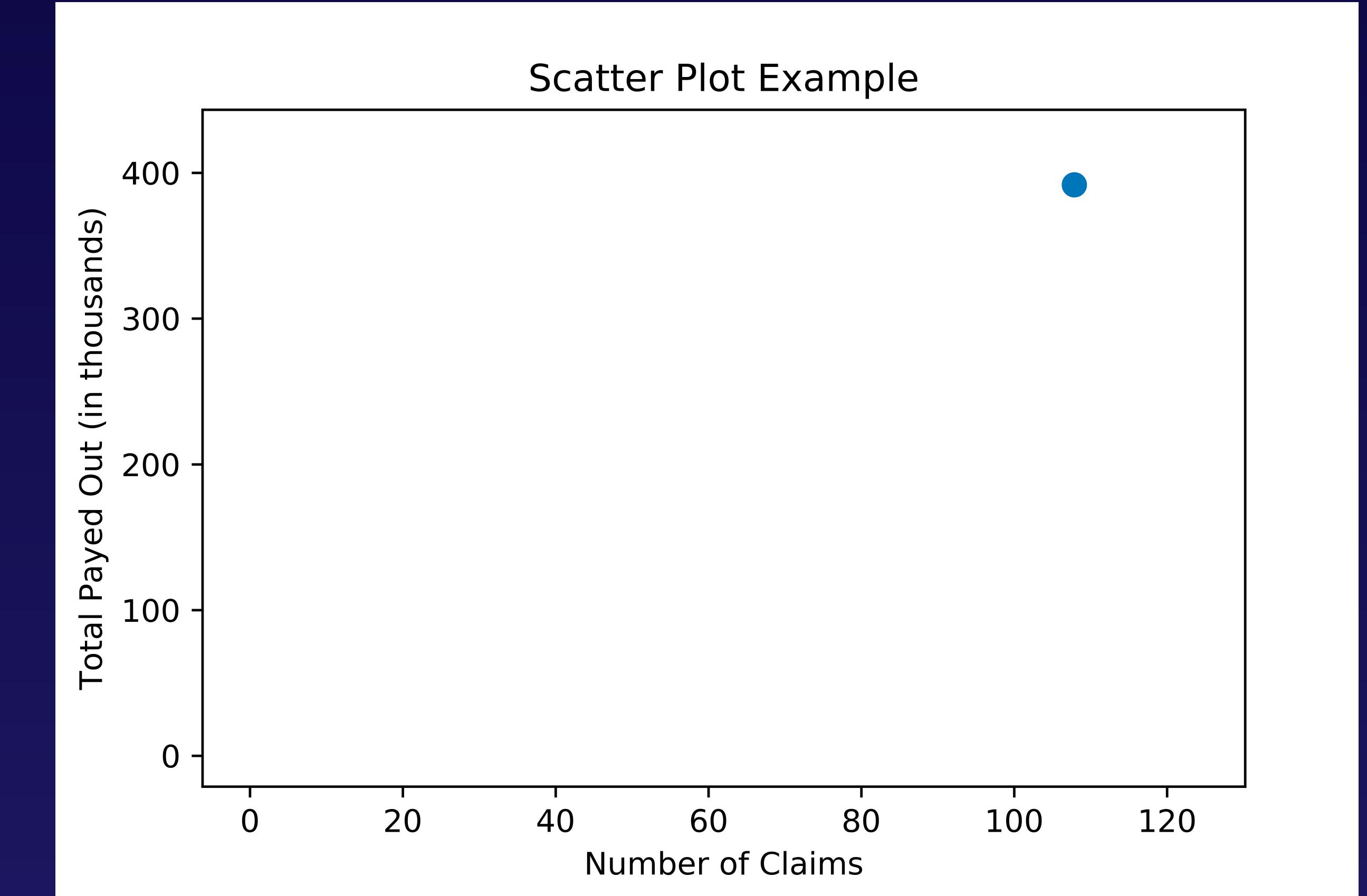
Scatter Plots

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



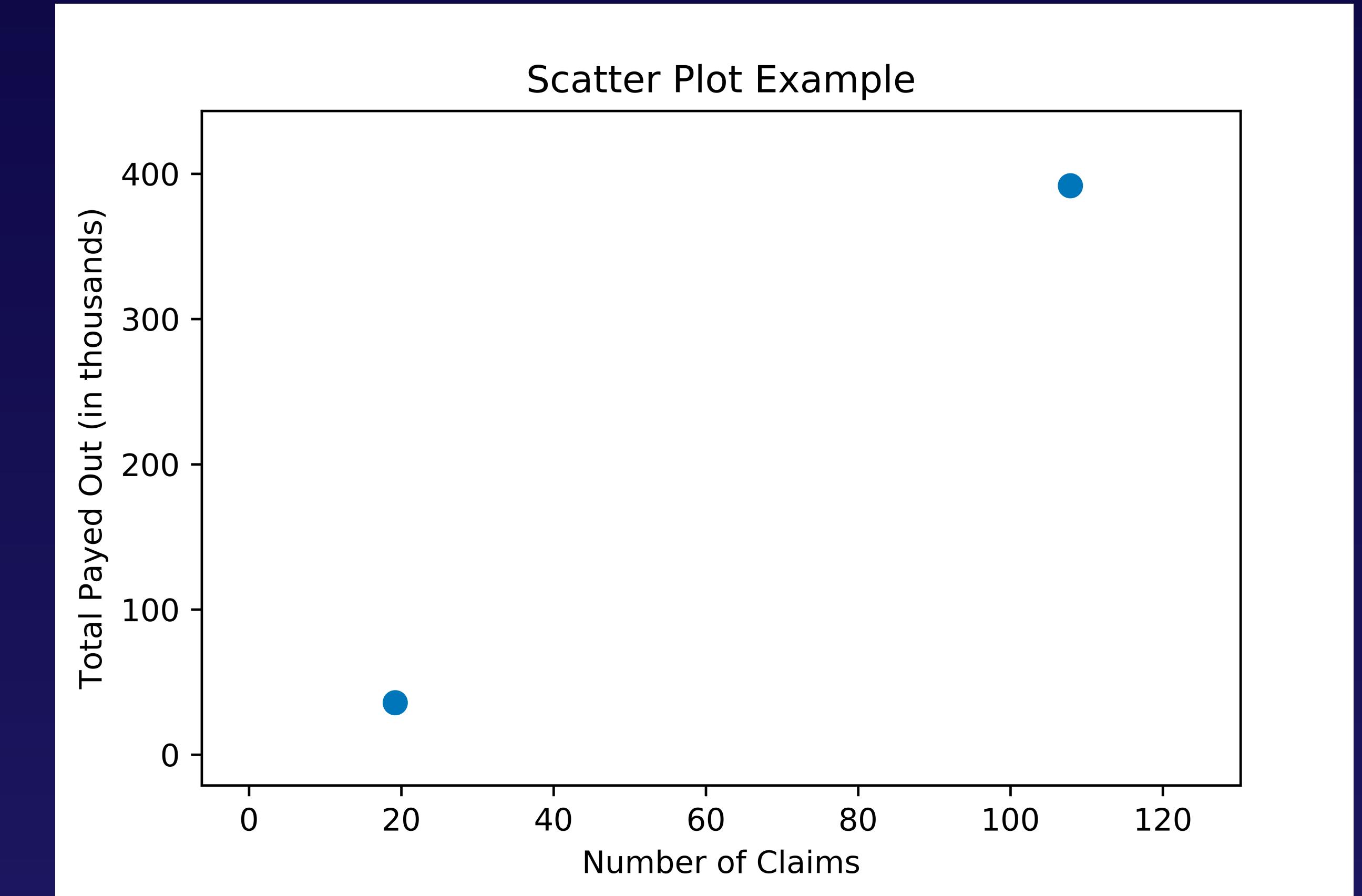
Scatter Plots

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



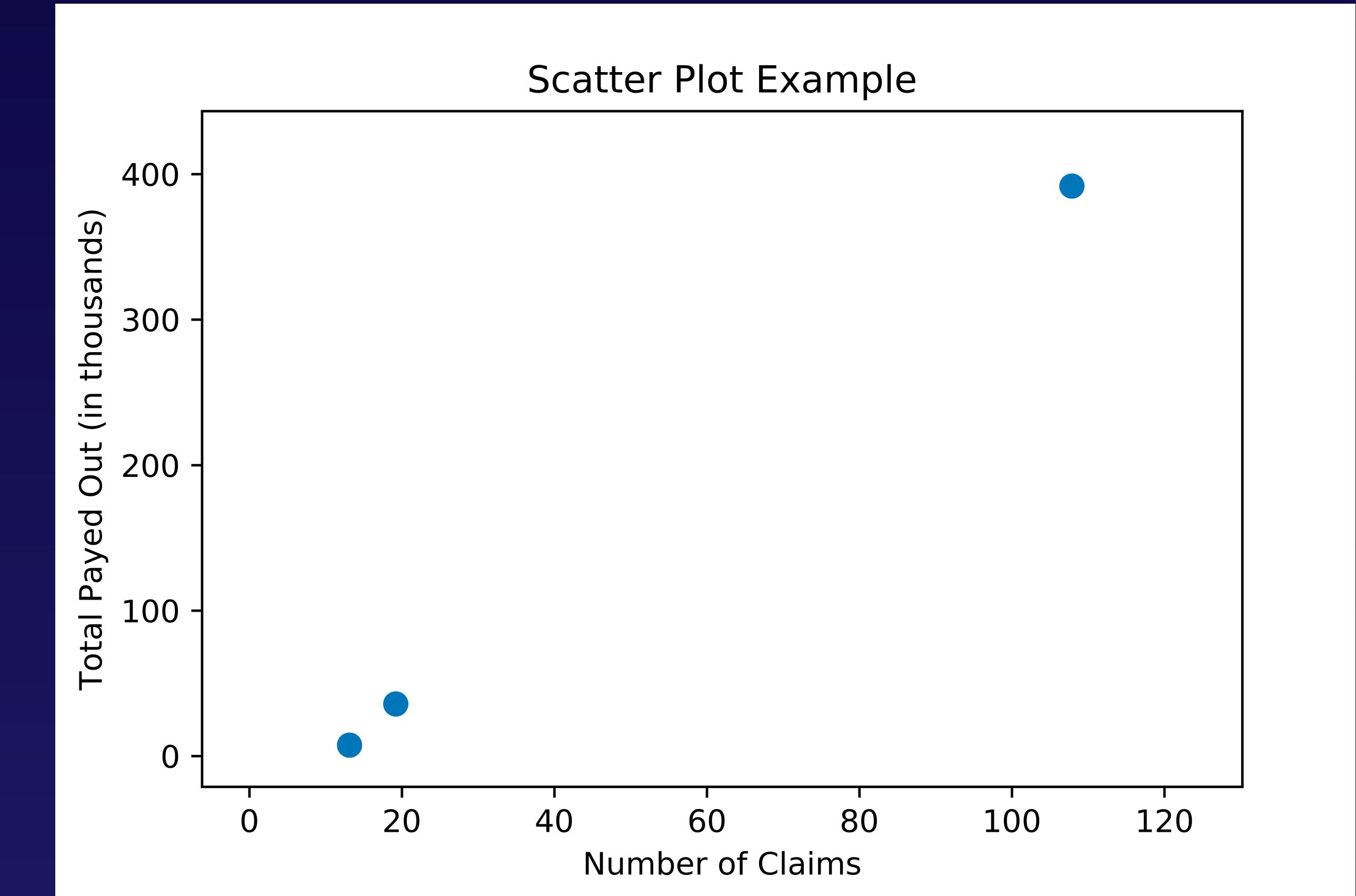
Scatter Plots

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



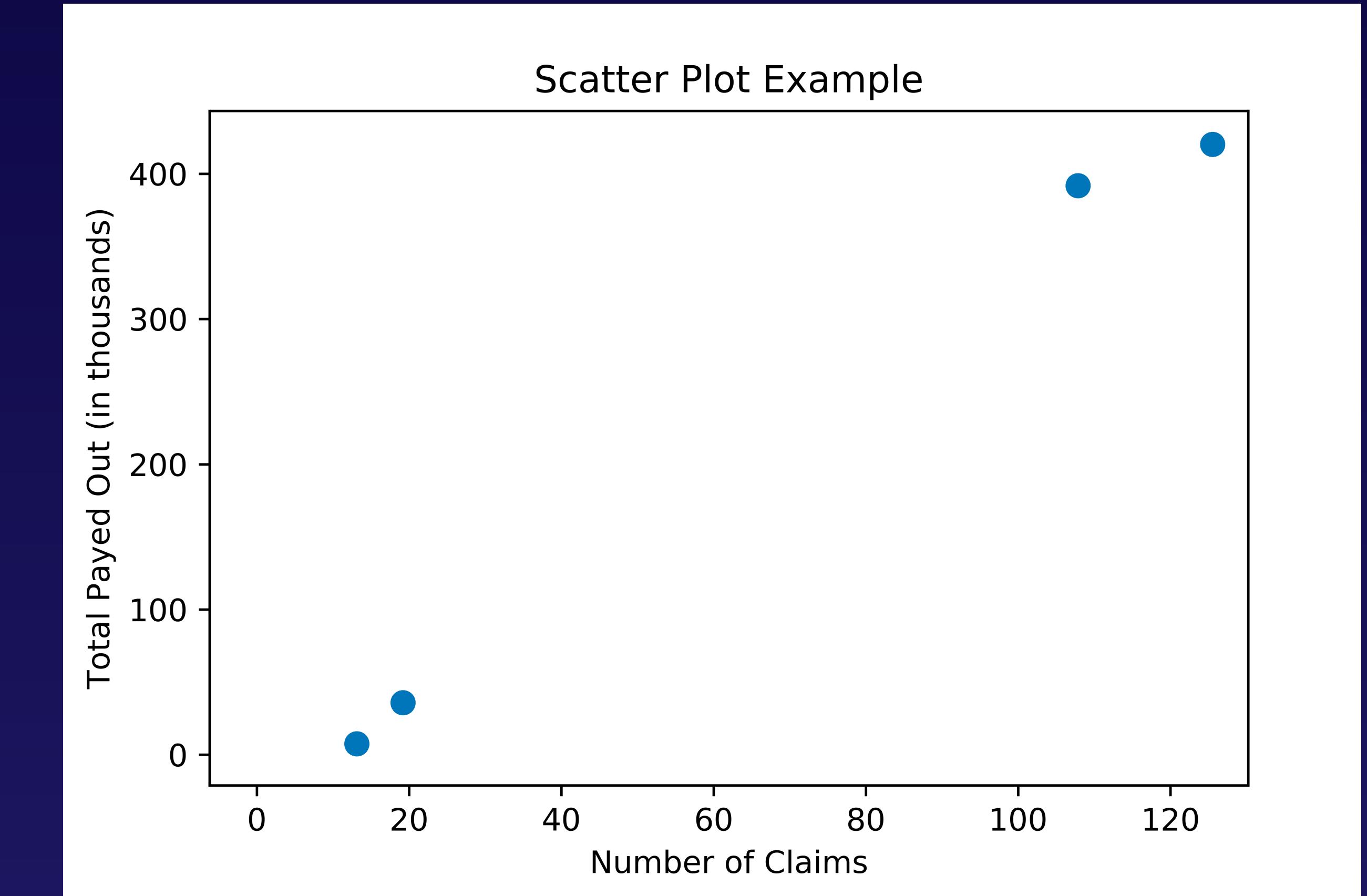
Scatter Plots

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



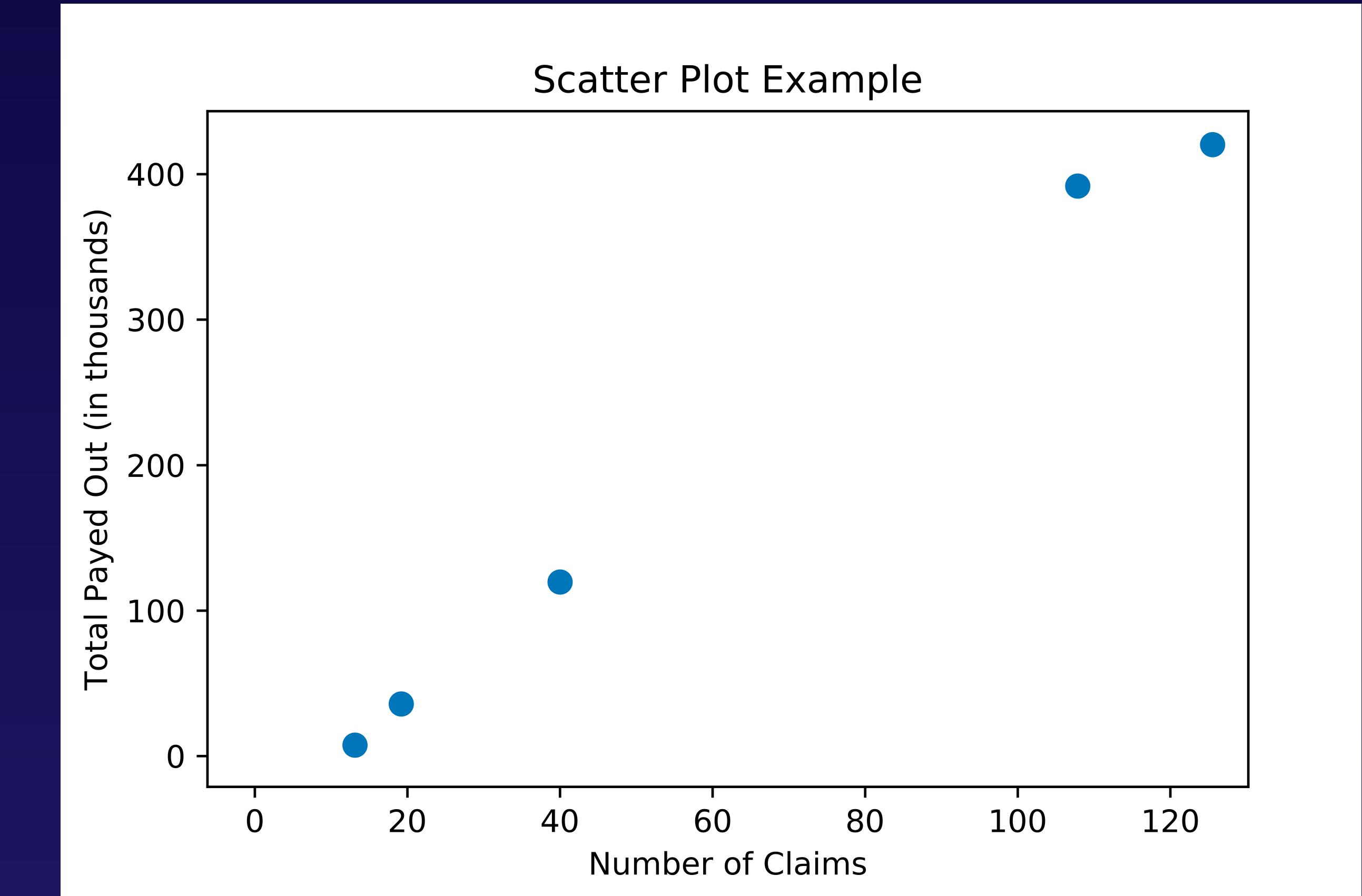
Scatter Plots

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



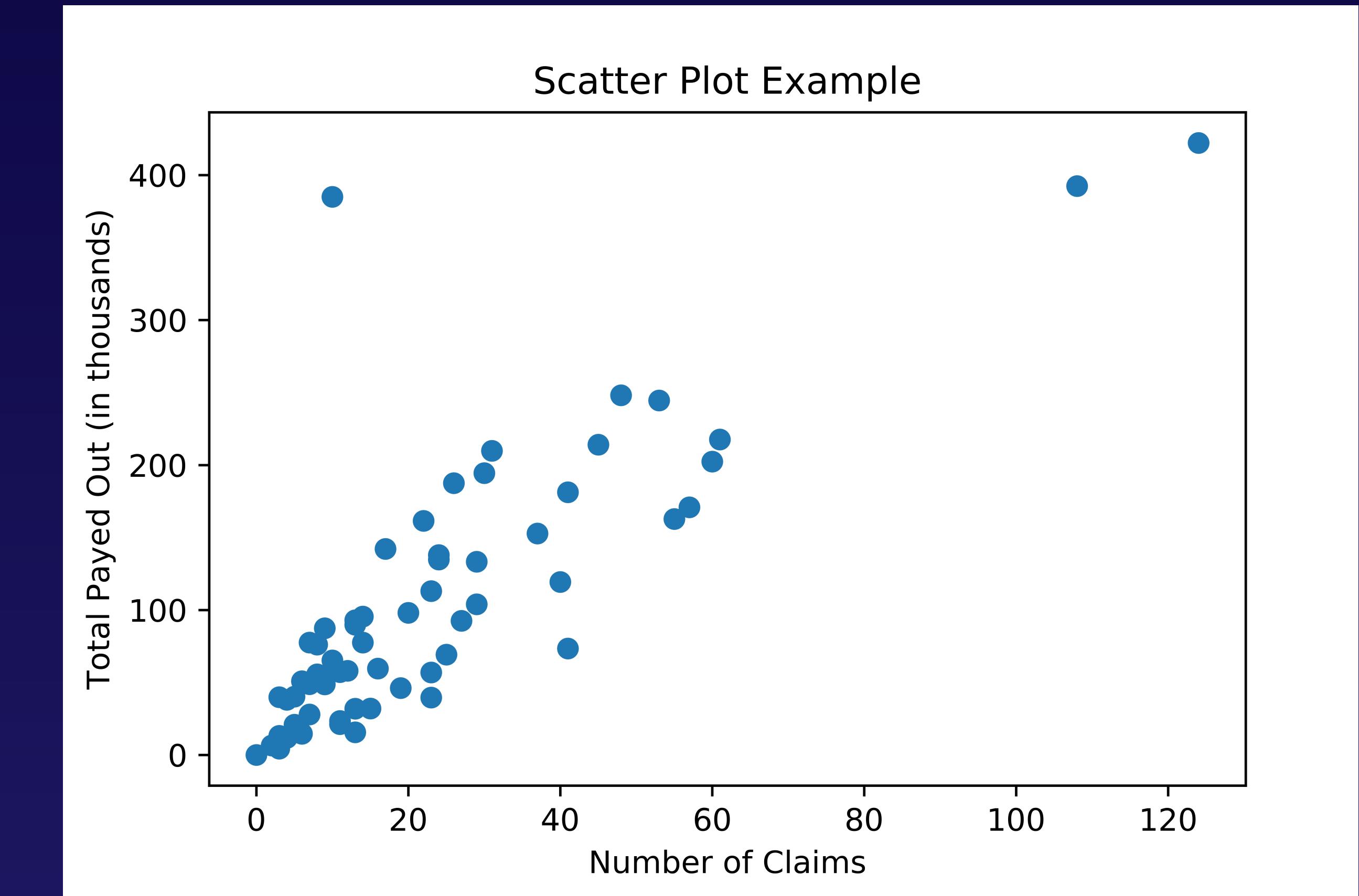
Scatter Plots

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



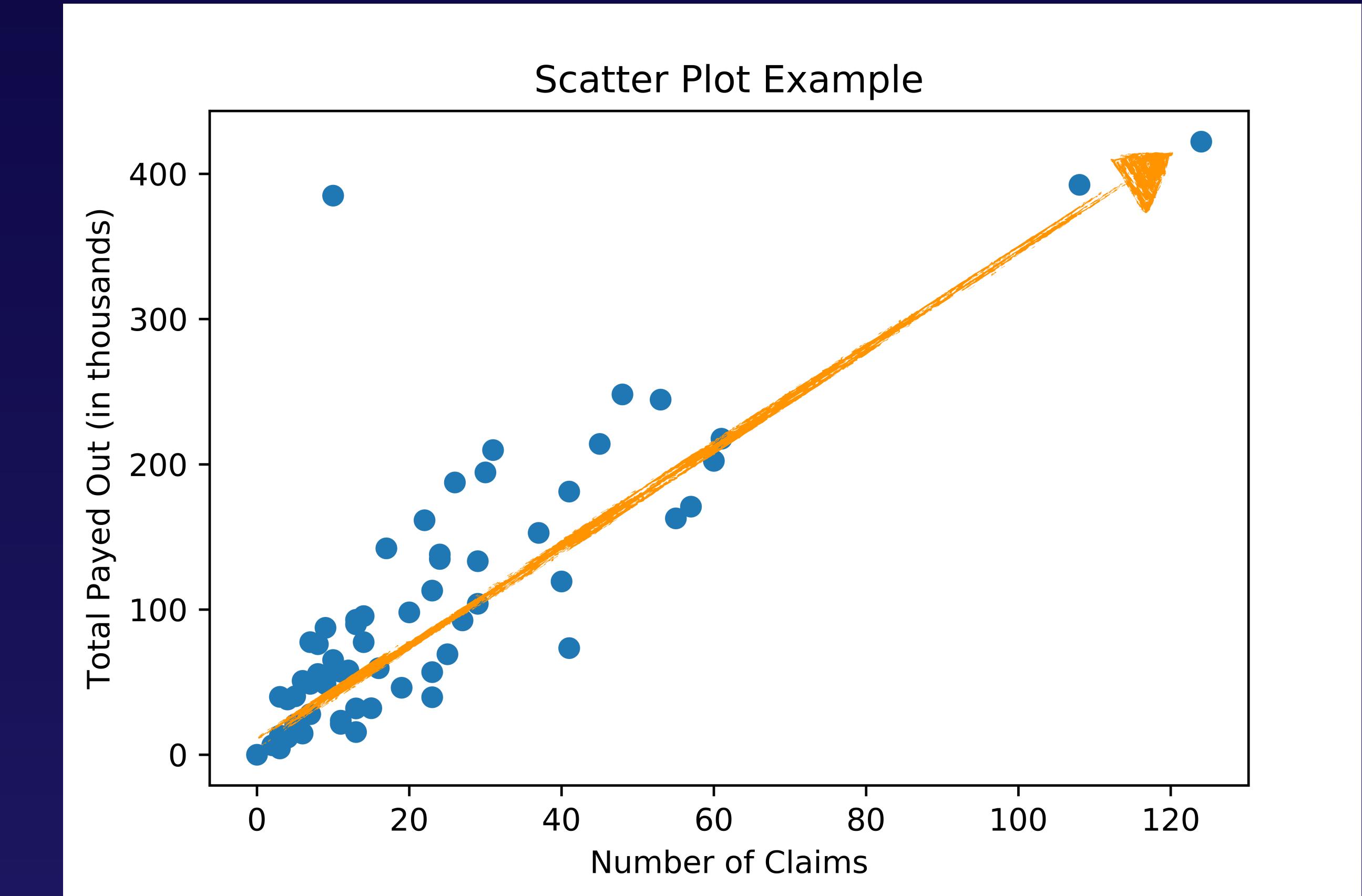
Scatter Plots

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



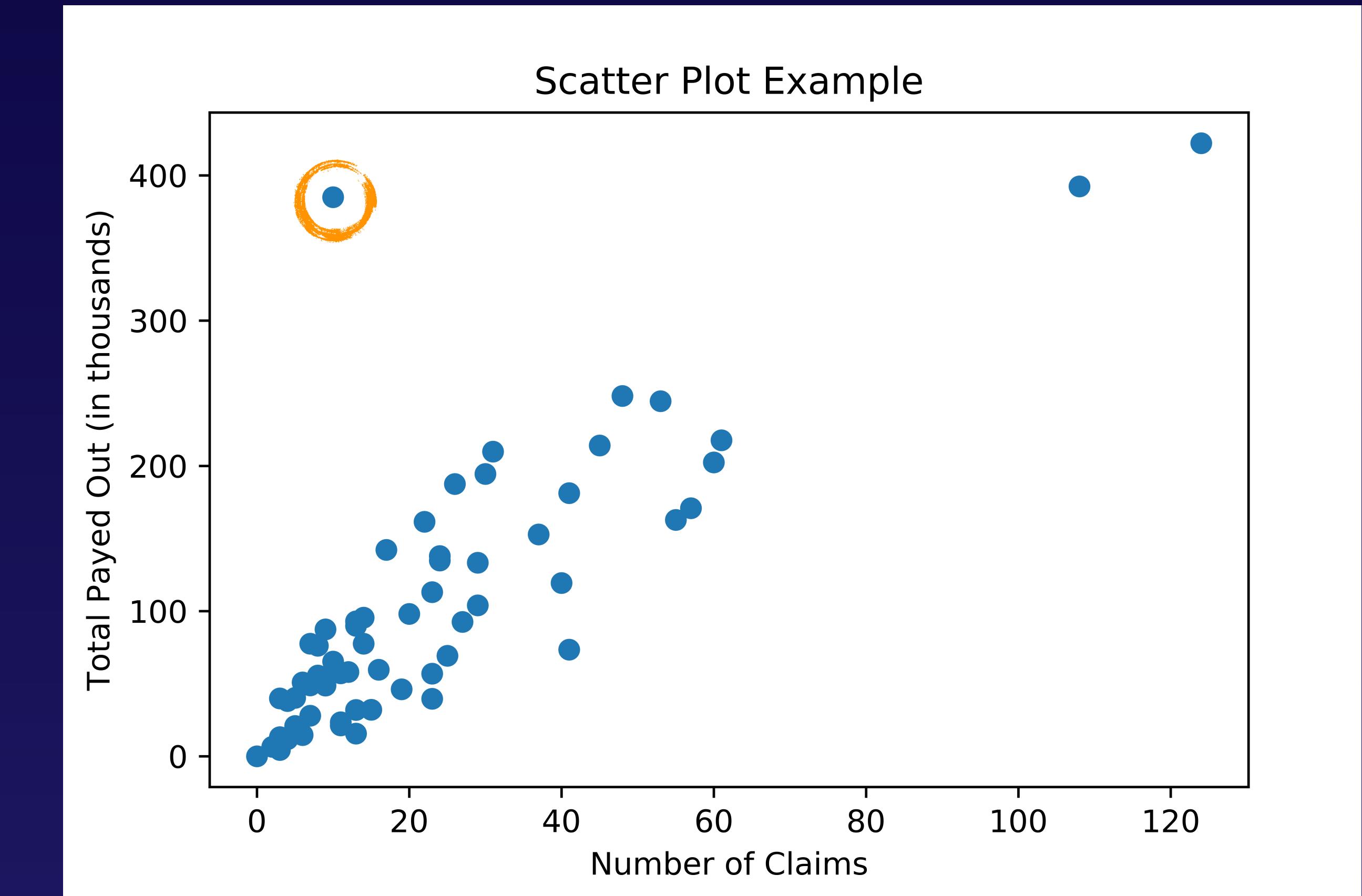
Scatter Plots

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



Scatter Plots

num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...



Bubble Plots

num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...

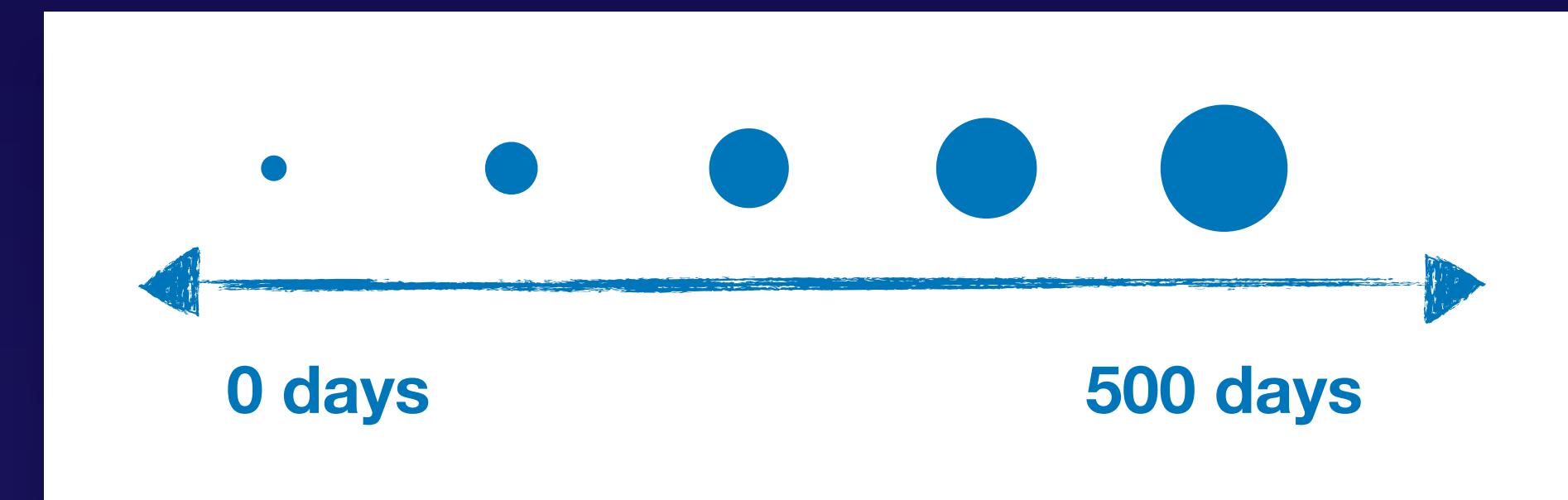
Bubble Plots

num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...

Bubble Plots

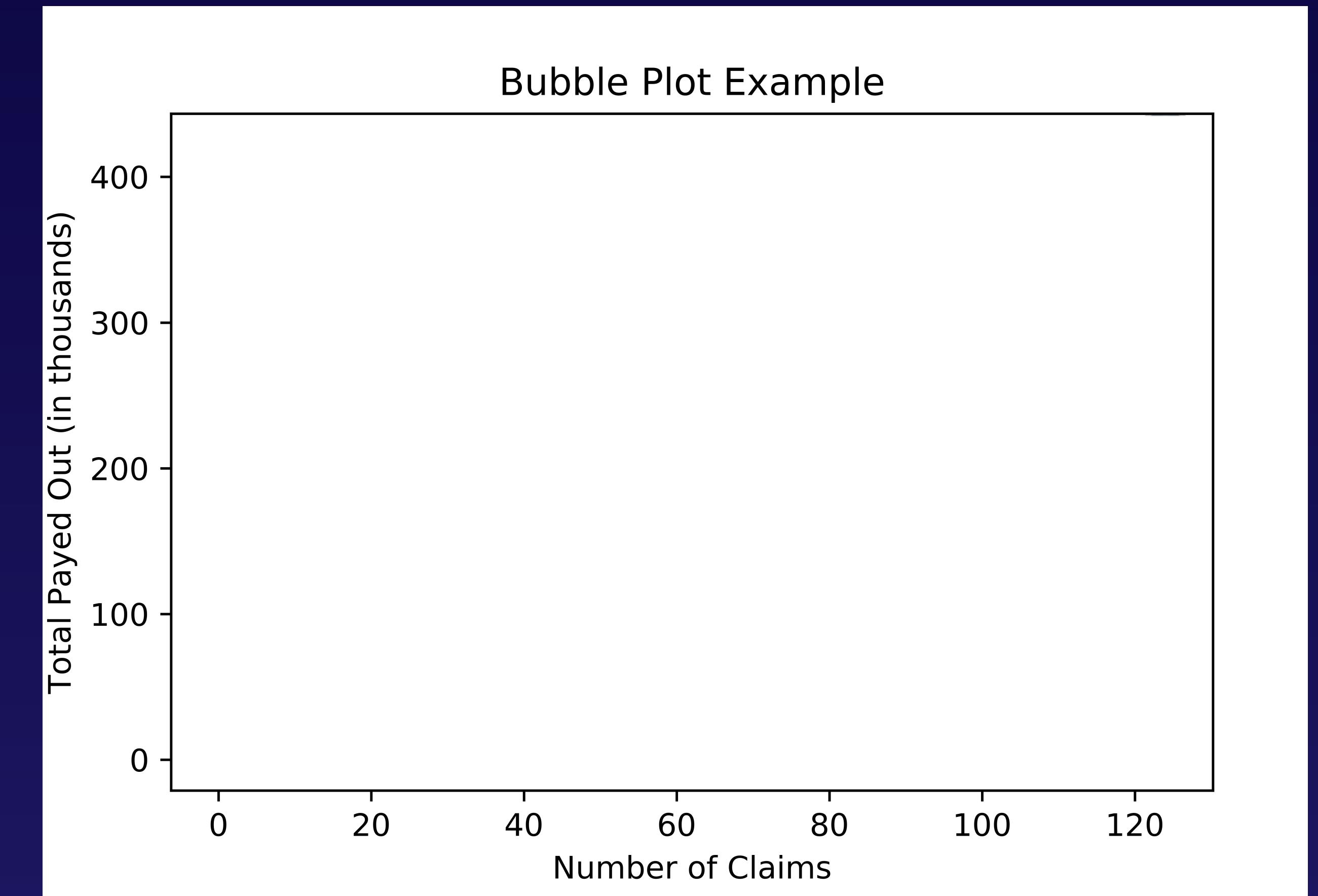
num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...

Bigger the bubble = more days



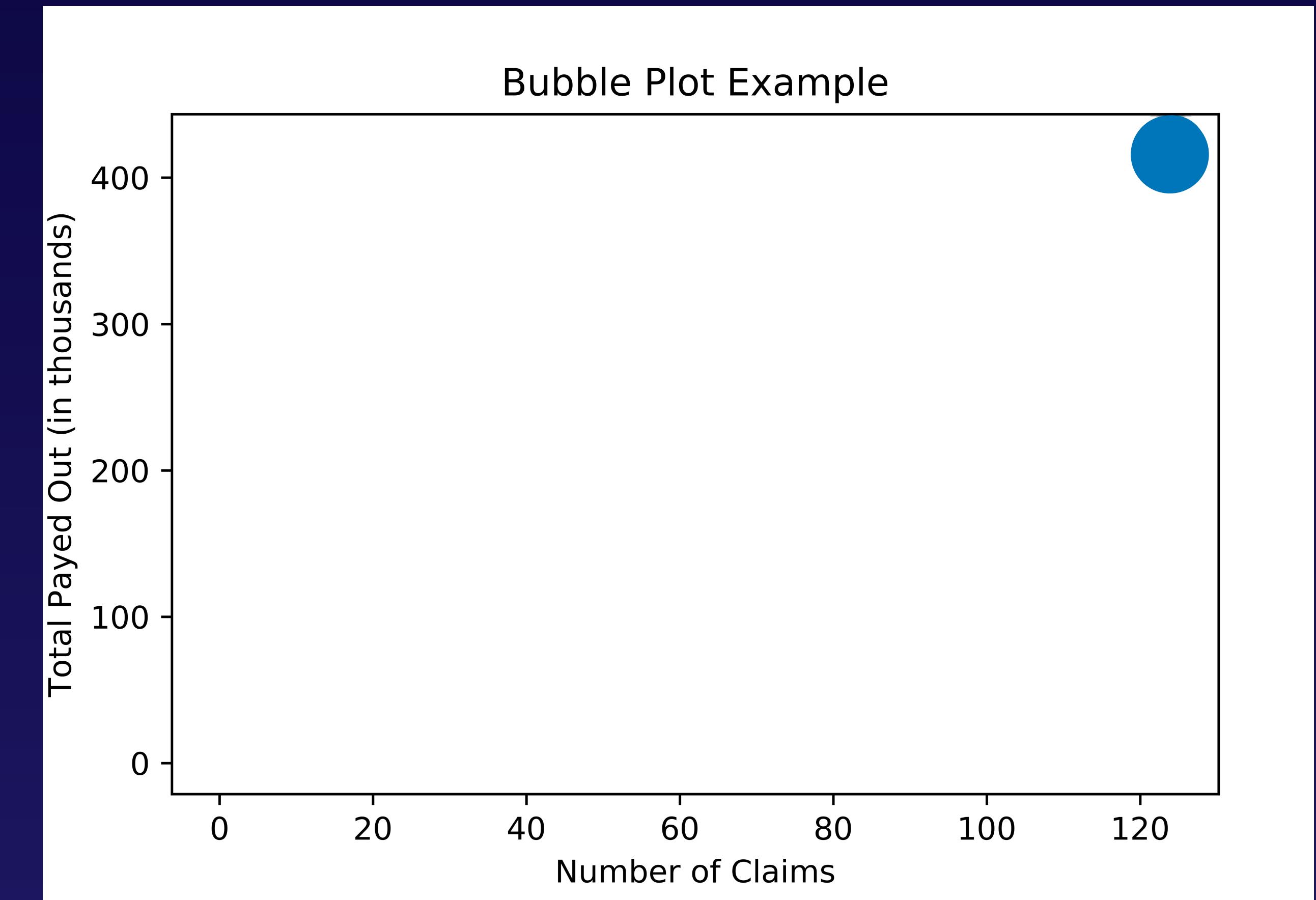
Bubble Plots

num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...



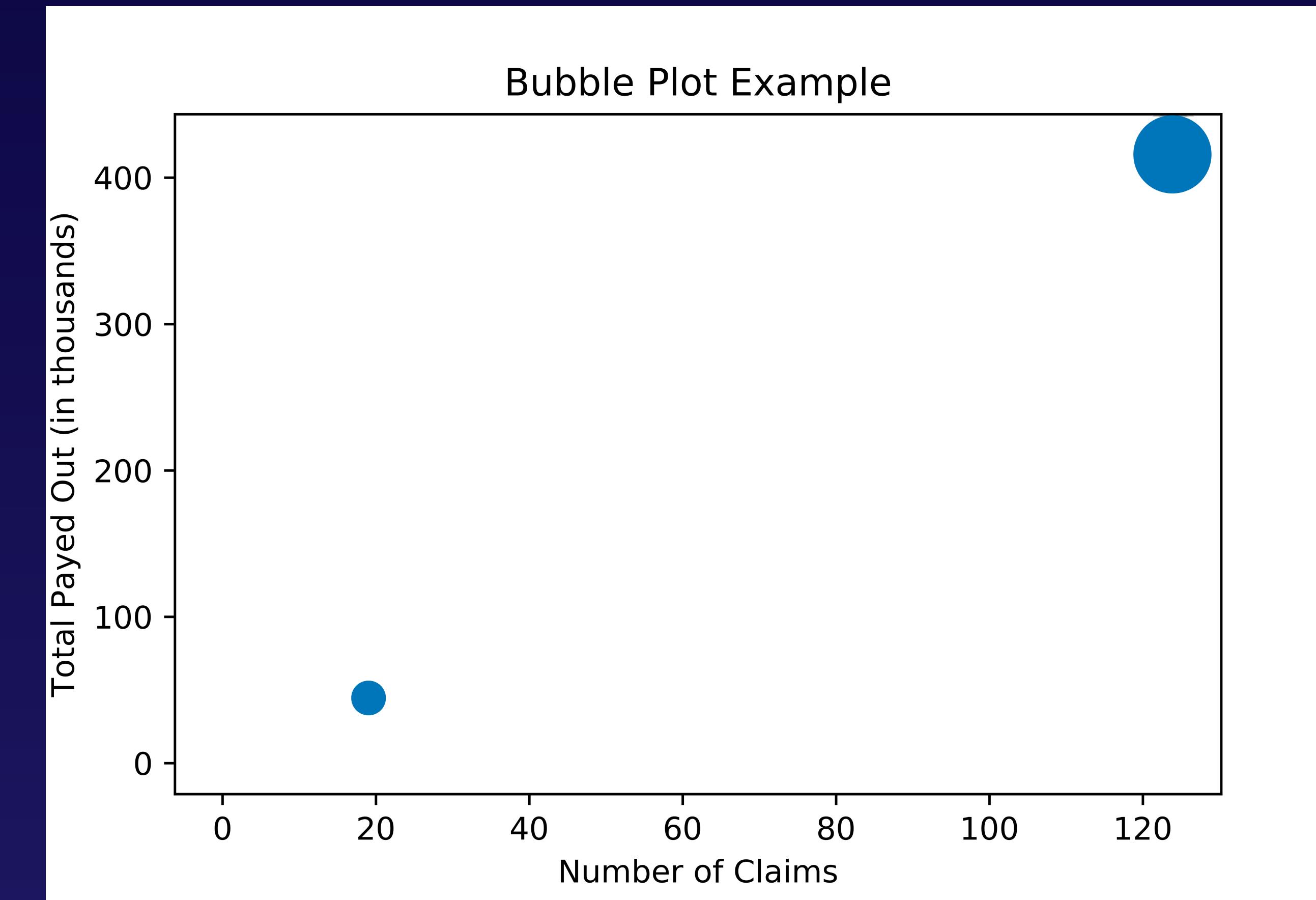
Bubble Plots

num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...



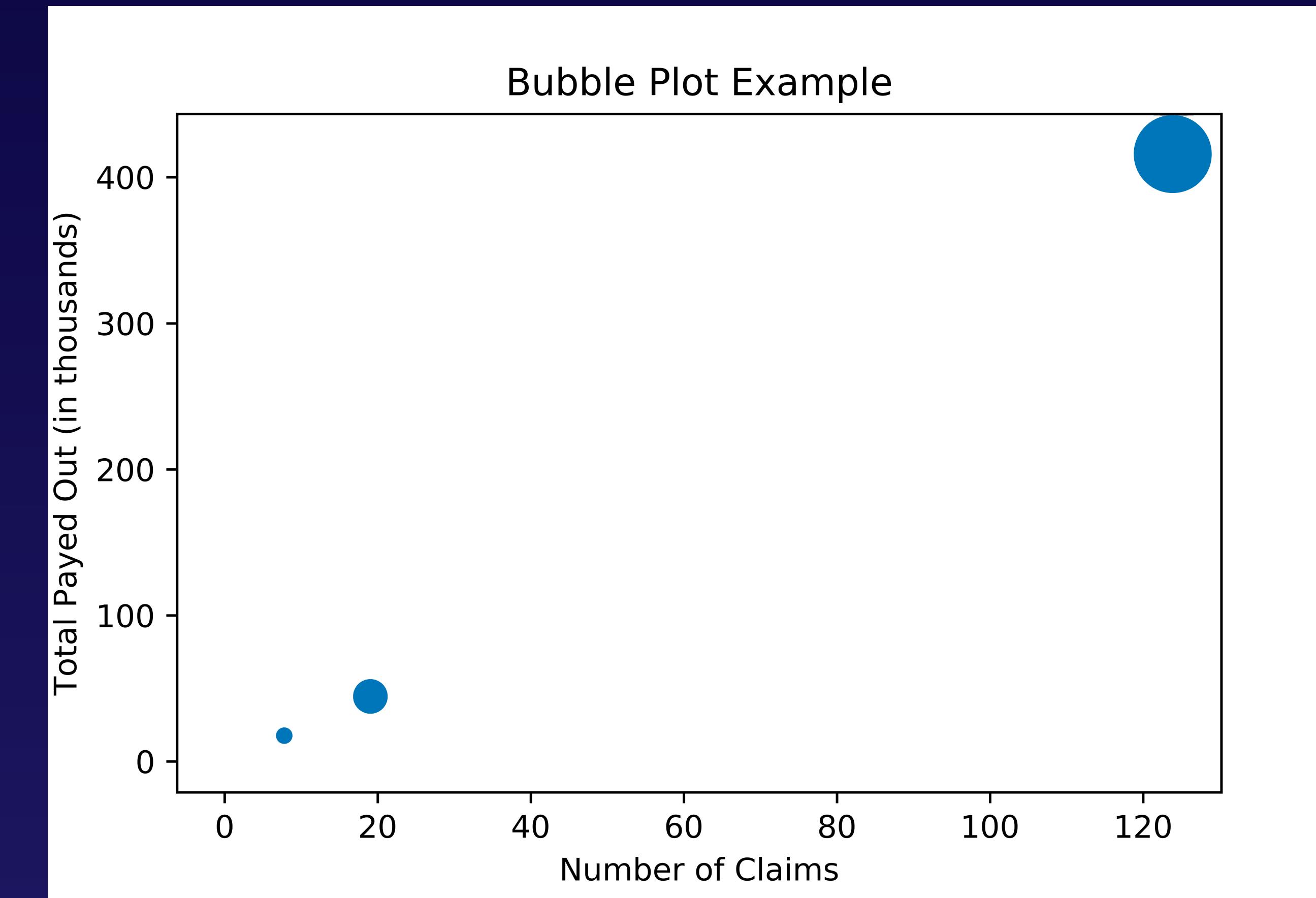
Bubble Plots

num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...



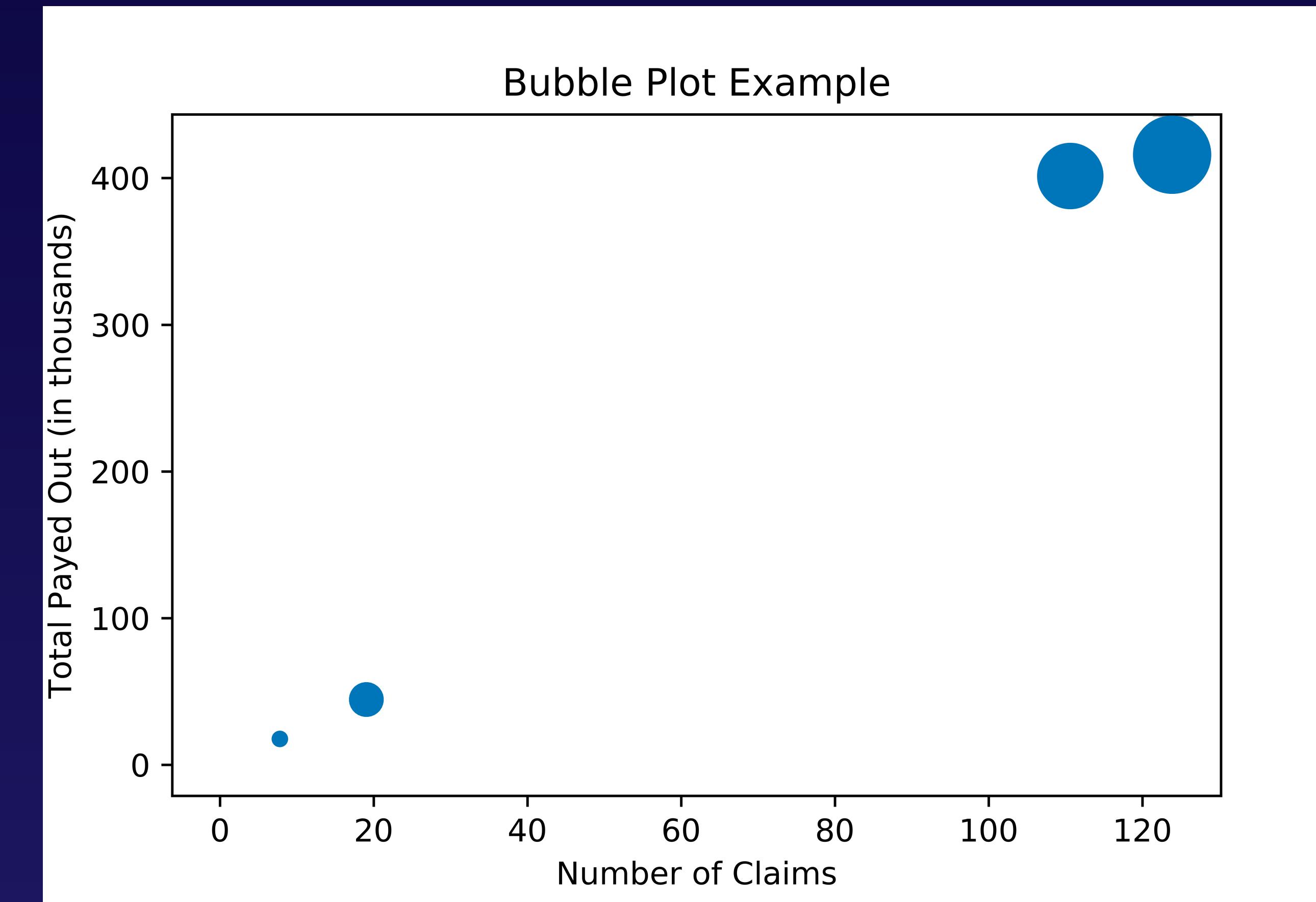
Bubble Plots

num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...



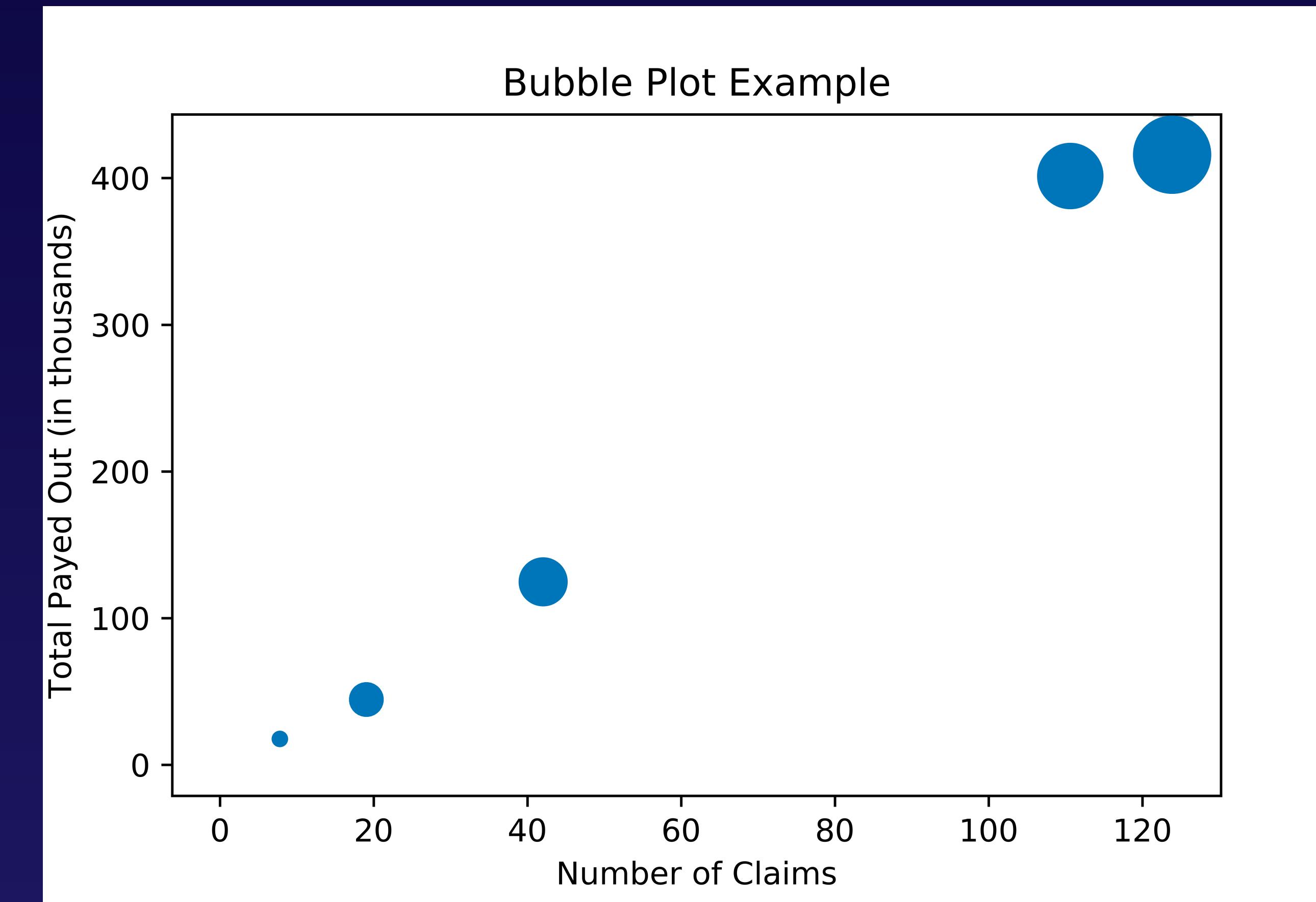
Bubble Plots

num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...



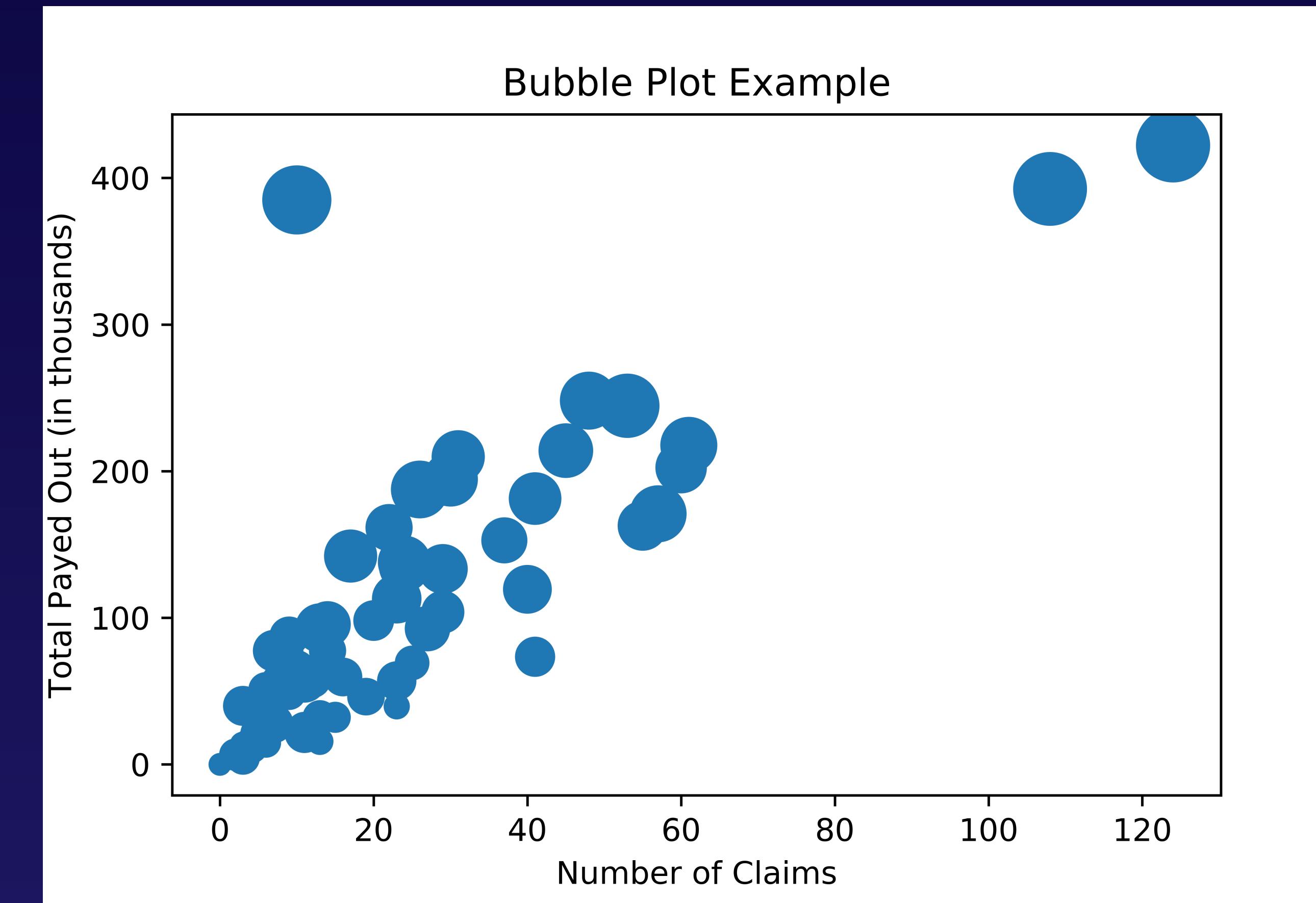
Bubble Plots

num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...

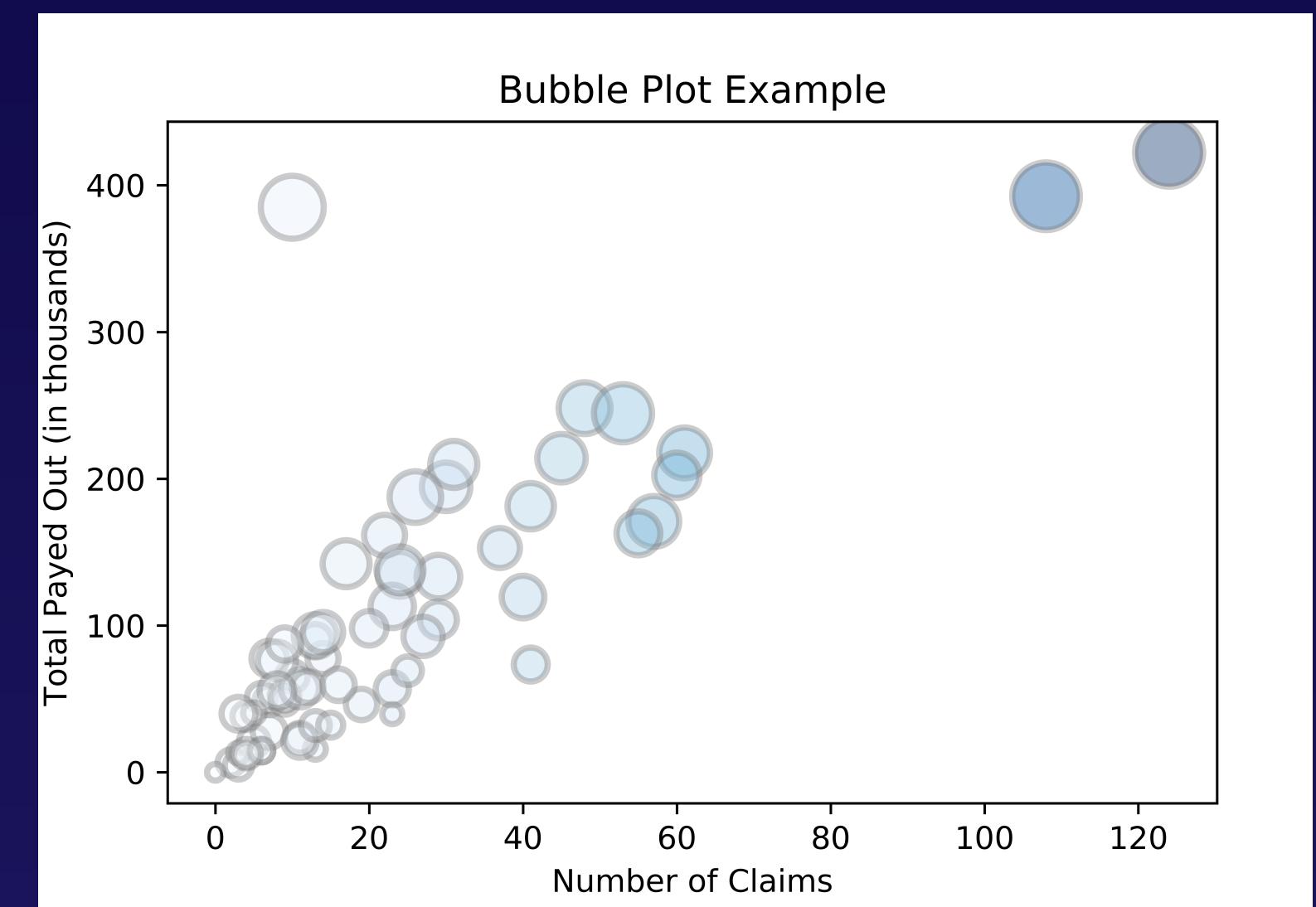
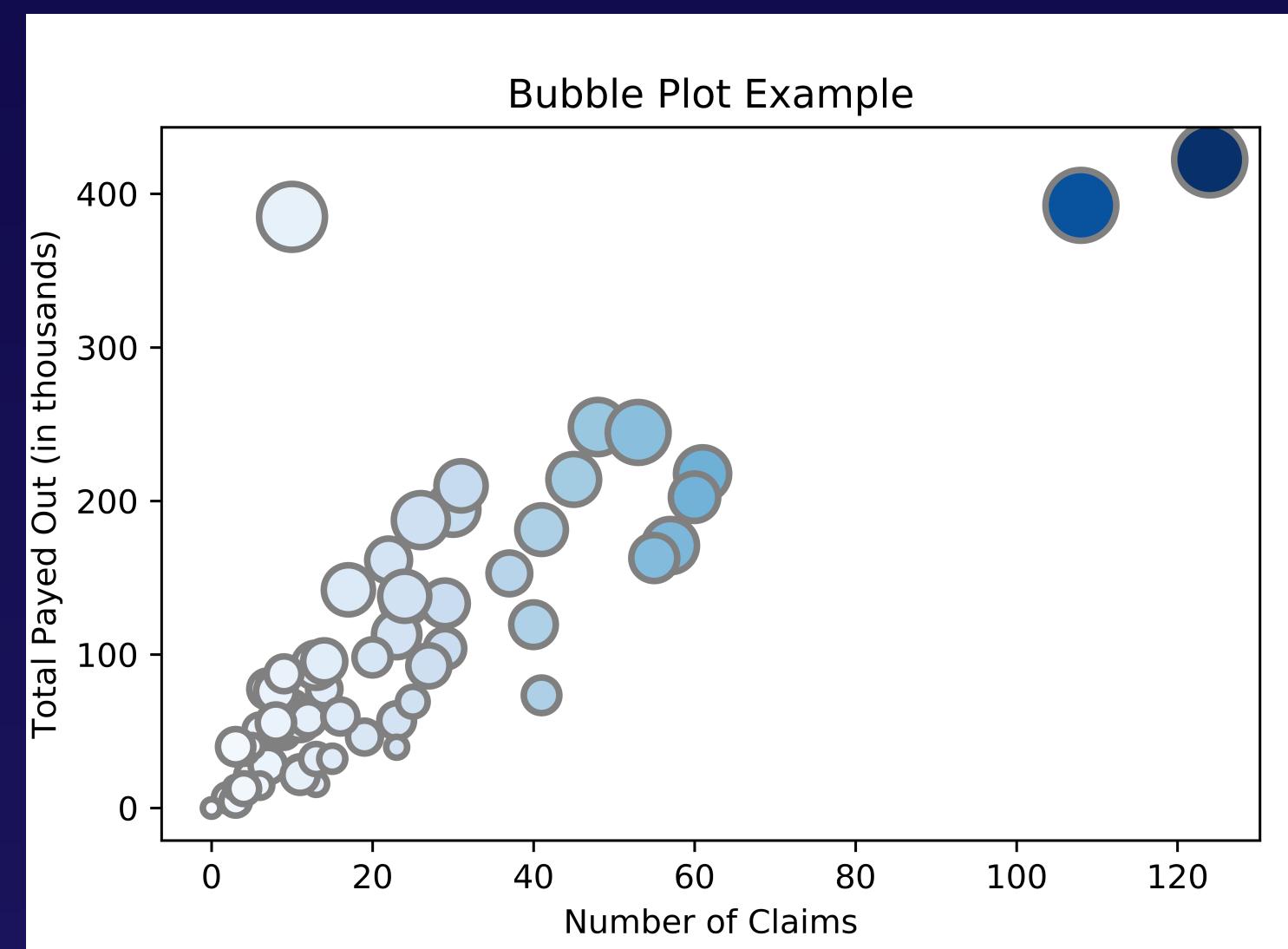
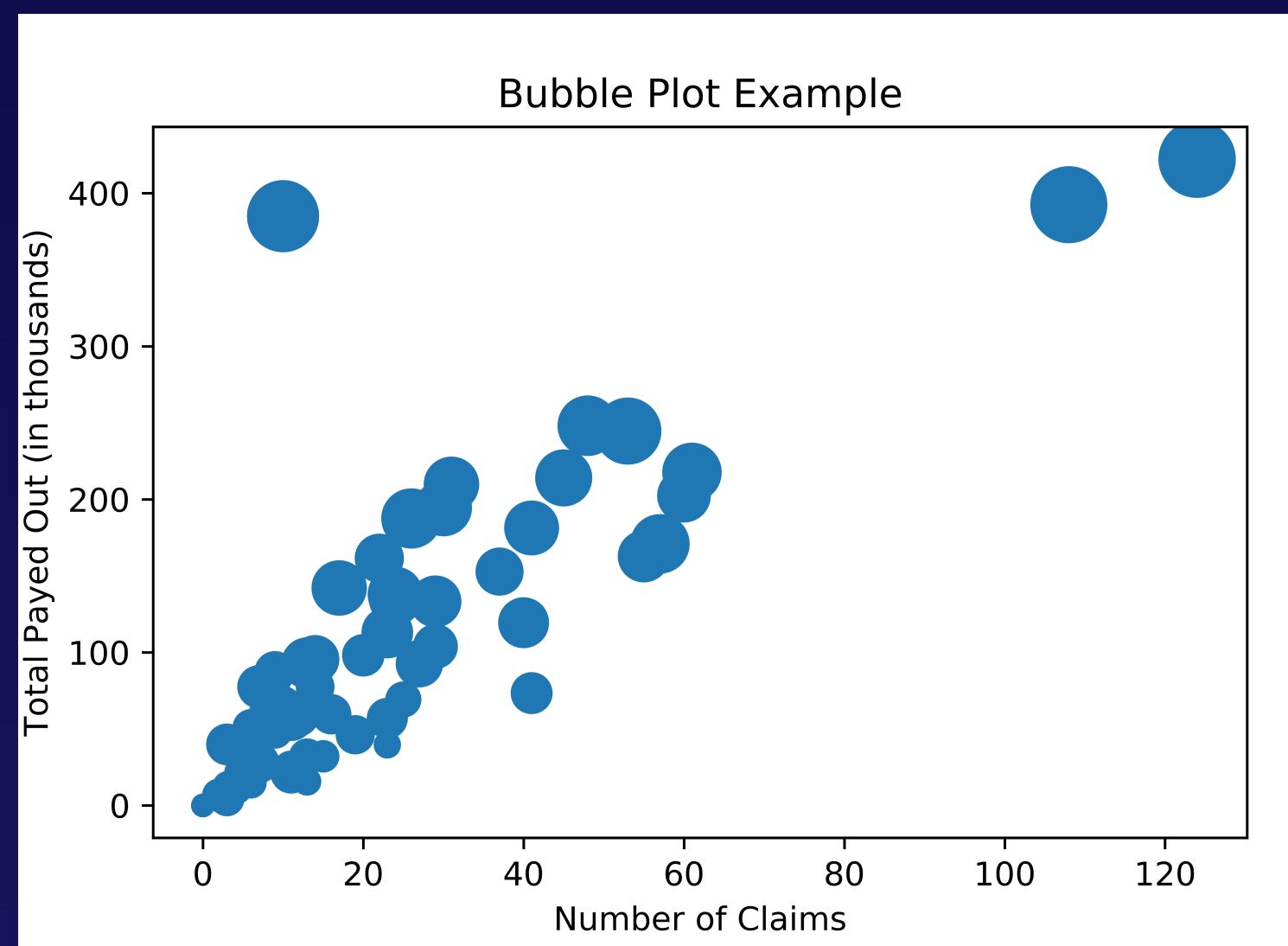


Bubble Plots

num_of_claims	total_payment	days_processed
108	392.5	480
19	46.2	41
13	15.7	5
124	422.2	350
40	119.4	54
...



Bubble Plots



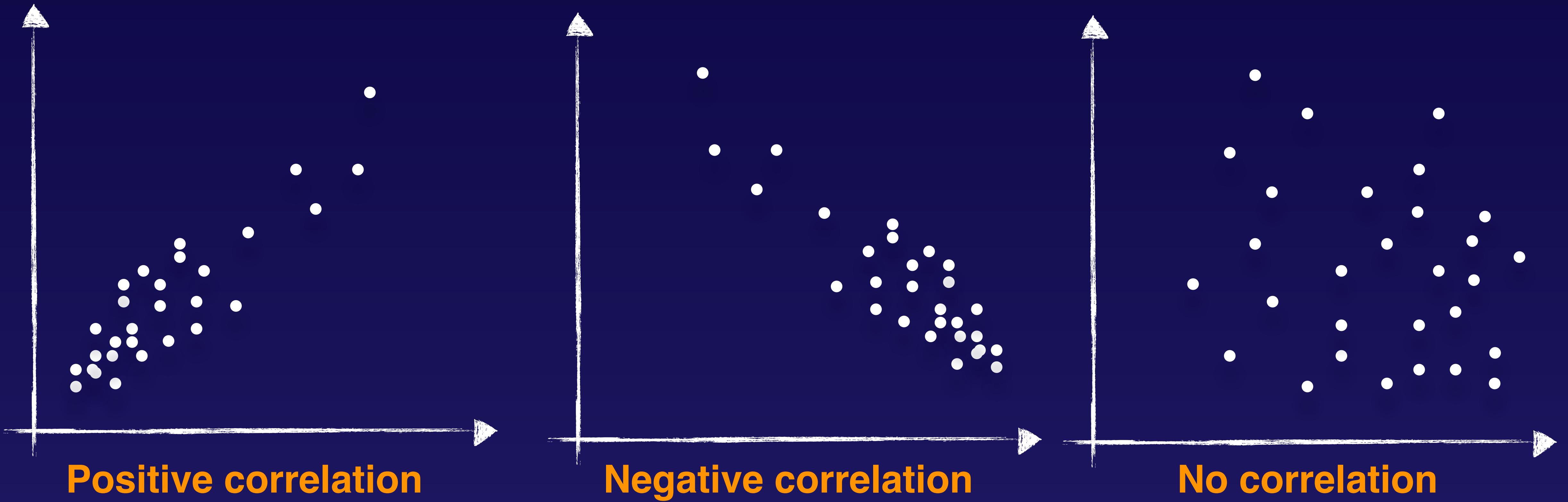
- **Scatter Plots**

Example: Is there any relationship between the size of a home and the price?

- **Bubble Plots**

Example: Are there any relationship between the size of a home, the age of the home, and the price?

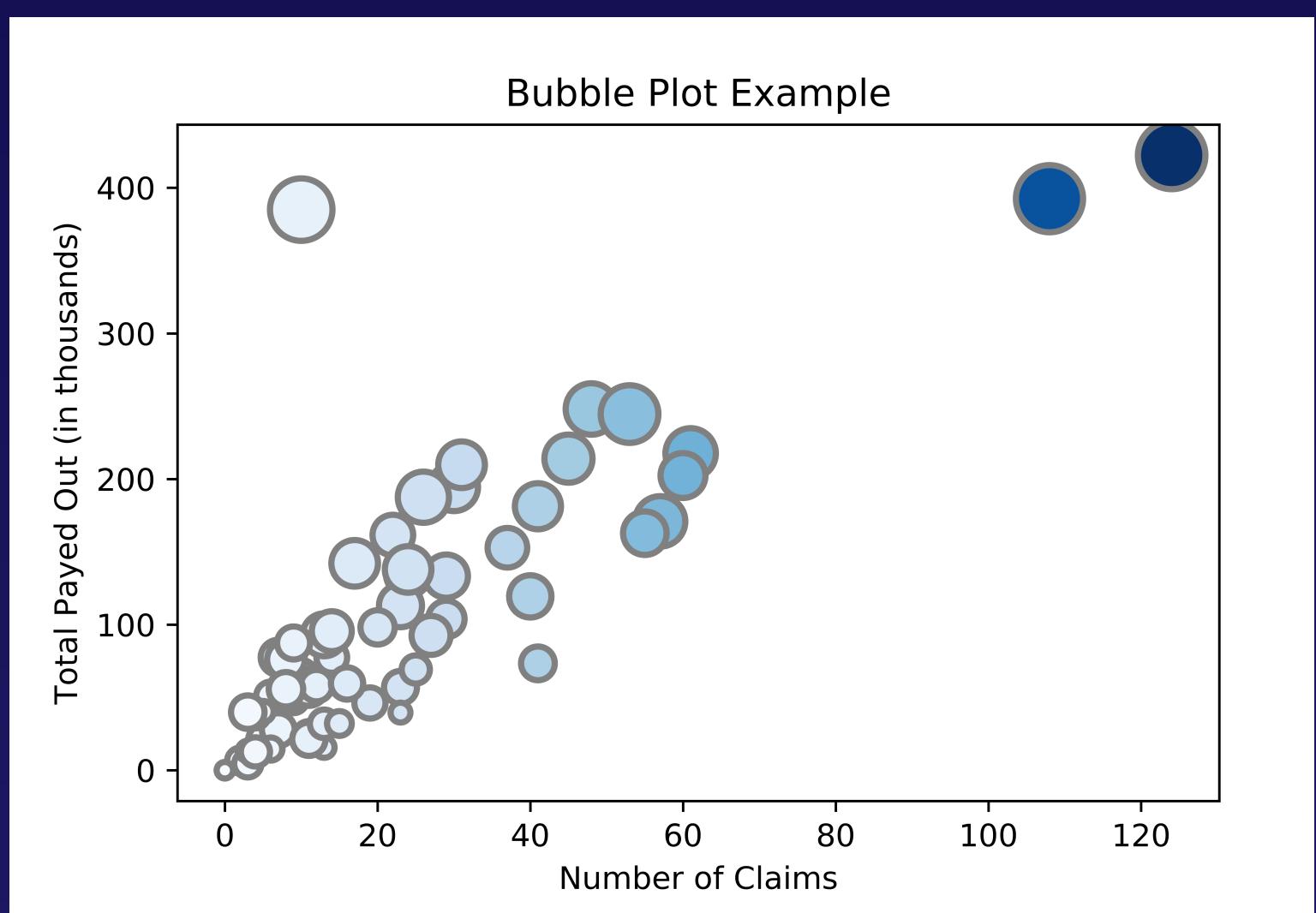
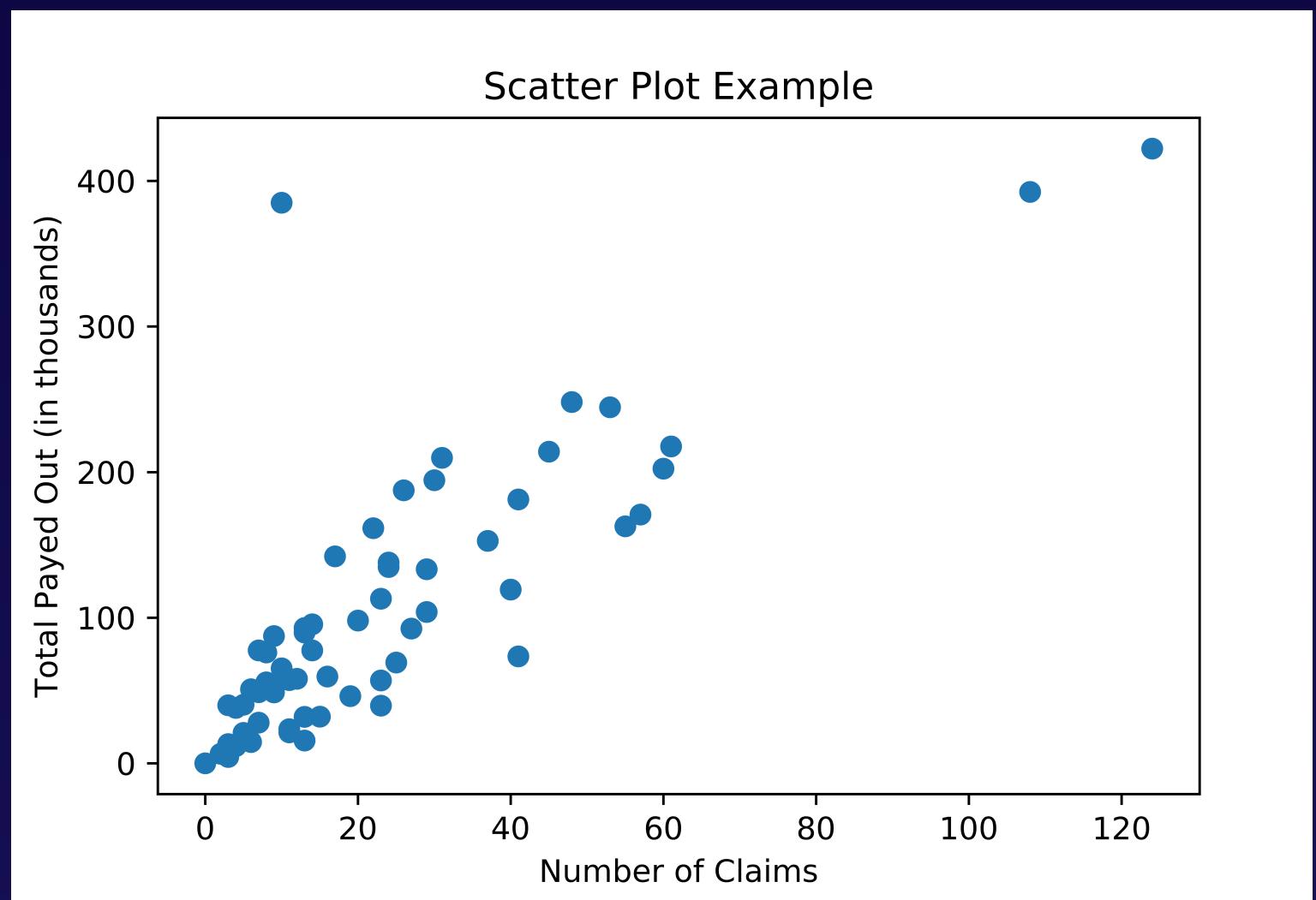
Correlation



Seeing Your Data - Relationships

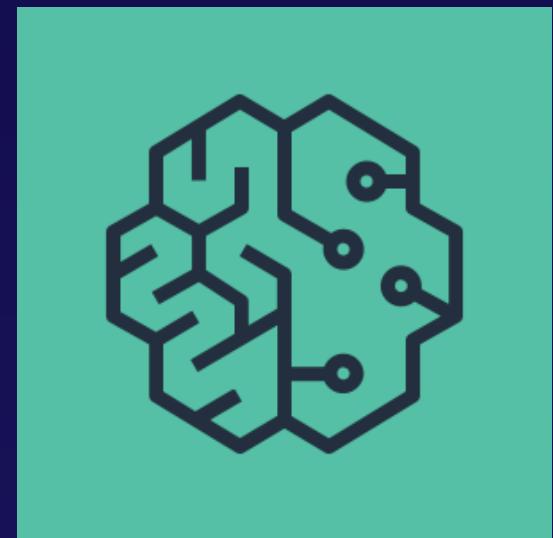
num_of_claims	total_payment
108	392.5
19	46.2
13	15.7
124	422.2
40	119.4
...	...

?



Seeing Your Data - Relationships

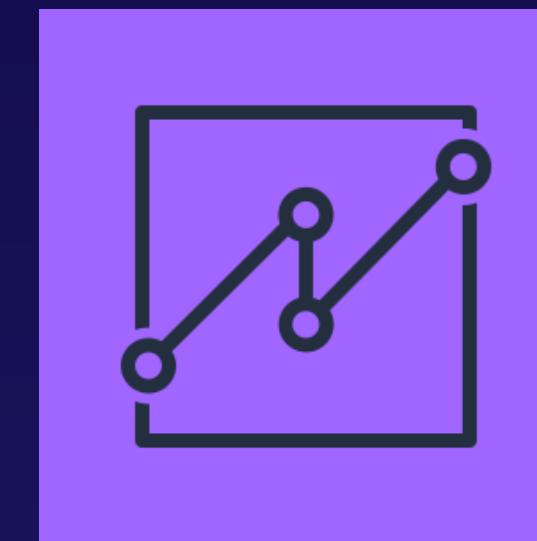
Developer Tools



SageMaker



Business Intelligence Tools (BI Tools)



Amazon QuickSight

Jupyter Notebooks

jupyter swedish-auto-scatter-plot-example Last Checkpoint: 02/28/2019 (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 2

In [30]: # Import different python libraries
%matplotlib inline
import sys
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

In [31]: # Read in data
df = pd.read_csv('./auto_insurance_data.csv')
df.head()

Out[31]:

	num_of_claims	total_payment
0	108	392.5
1	19	46.2
2	13	15.7
3	124	422.2
4	40	119.4

In [32]: # Plot graph
plt.scatter(x=df['num_of_claims'],y=df['total_payment'])
plt.title('Scatter Plot Example')
plt.xlabel('Number of Claims')
plt.ylabel('Total Payed Out (in thousands)')
plt.savefig('./images/Scatter_Plot_Example.png', format='png', dpi=1200)
plt.show()

Scatter Plot Example

