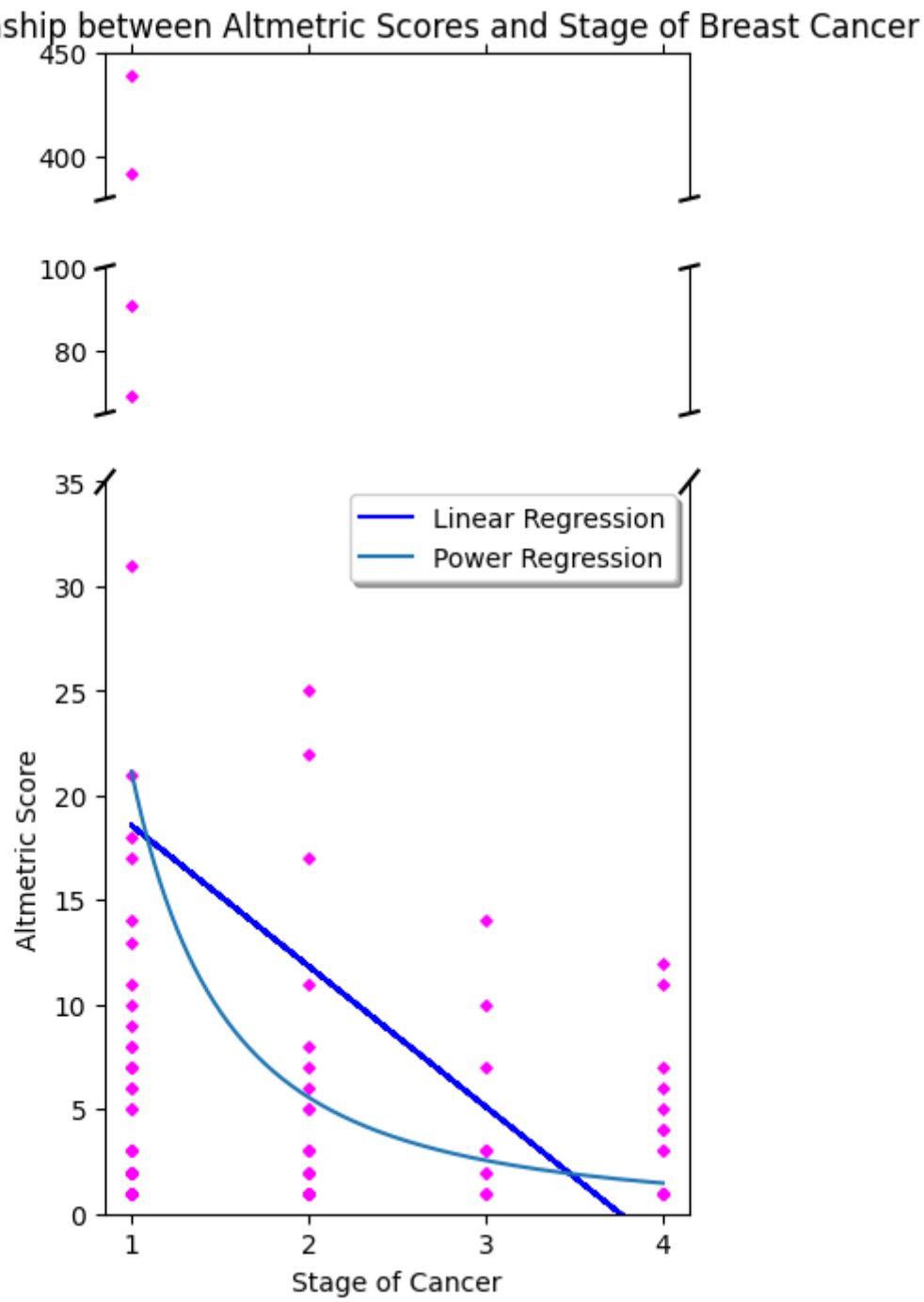
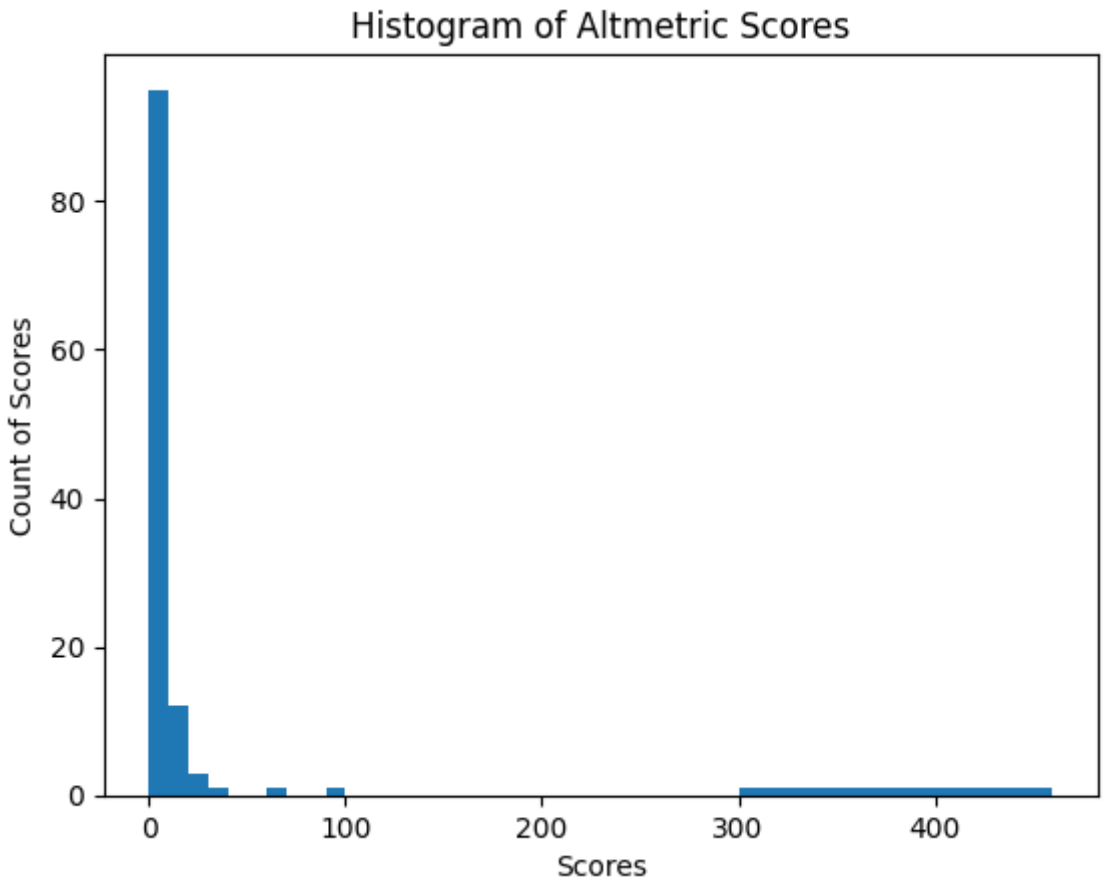
In [104…

```
# Relation Between The Altmetric Score and Severity of Cancer
# Programmers: Ben Ford, Brady Balk, Alexander Al-Taraireh
# The purpose of this graph is to demonstrate and present the relationship betw
# DATA SOURCE: https://www.dropbox.com/s/8jjzdcbtgyx9lht/Dimensions-Clinical-Tr
# Libraries:
import matplotlib.pyplot as plt
import numpy as np
from scipy.stats import linregress
from scipy.optimize import curve_fit
# Creating the data:
# x represents the Stage of Cancer
# y represents the Altmetric Score
x = np.array([2,3,1,1,2,4,1,1,1,2,2,1,2,2,2,2,2,1,1,1,1,1,1,1,1,1,2,4,4,4,4,2,2
y = np.array([1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1
# Get linear equation
(slope, intercept, rvalue, pvalue, stderr) = linregress(x,y)
# Power regression equation
popt, pcov = curve_fit(lambda fx,a,b: a*fx**-b,  x,  y)
x_linspace = np.linspace(min(x), max(x), 456)
power_y = popt[0]*x_linspace**-popt[1]
perr = np.sqrt(np.diag(pcov))
# Changing shape with marker
f, (ax,ax2,ax3) = plt.subplots(3,1,sharex=True, figsize=(4, 8), gridspec_kw={'h
ax.scatter(x, y, s=8, color=[1,0,1], marker="D")
ax.title.set_text('Relationship between Altmetric Scores and Stage of Breast Ca
ax2.scatter(x, y, s=8, color=[1,0,1], marker="D")
ax3.scatter(x, y, s=8, color=[1,0,1], marker="D")
ax.set_ylim(380, 450)
ax2.set_ylim(65,100)
ax3.set_ylim(0, 35)
# Hiding the spines between ax, ax2, and ax3
ax.spines['bottom'].set_visible(False)
ax2.spines['top'].set_visible(False)
ax2.spines['bottom'].set_visible(False)
ax2.tick_params(bottom=False)
ax.xaxis.tick_top()
ax.tick_params(labeltop=False)
ax3.spines['top'].set_visible(False)
# Will show how large to make the diagnal lines in the axes coordinates
d = .015
# Arguments to pass to plot, so that they wont be repeated
kwargs = dict(transform=ax.transAxes, color='k', clip_on=False)
# Top-left diagonal
ax.plot((-d, +d), (-d, +d), **kwargs)
# Top-right diagonal
ax.plot((1 -d, 1 + d), (-d, +d), **kwargs)
# Switch to the bottom axes
kwargs.update(transform=ax2.transAxes)
# Bottom-left diagonal
ax2.plot((-d, +d), (1 -d, 1 + d), **kwargs)
# Bottom-right diagonal
ax2.plot((1 -d, 1 + d), (1 -d, 1 + d), **kwargs)
# Top-left diagonal 2
ax2.plot((-d, +d), (-d, +d), **kwargs)
# Top-right diagonal 2
ax2.plot((1 -d, 1 + d), (-d, +d), **kwargs)
```

```
# Switch to the bottom axes
kwargs.update(transform=ax3.transAxes)
# Bottom-left diagonal
ax3.plot((-d, +d), (1 -d, 1 + d), **kwargs)
# Bottom-right diagonal
ax3.plot((1 -d, 1 + d), (1 -d, 1 + d), **kwargs)
# Tick marks for x axis
plt.xticks([1,2,3,4])
# Linear Regression calcuation
y_pred = intercept + slope*x
# Plot linear regression
plt.plot(x,y_pred, color="blue", label="Linear Regression")
# Power regression
plt.plot(x_linspace, power_y, label='Power Regression')
# Labels for x and y axes
plt.xlabel('Stage of Cancer')
plt.ylabel('Altmetric Score')
# Show the graph
plt.legend(loc='best', fancybox=True, shadow=True)
plt.show()
plt.hist(y, bins = [0,10,20,30,40,50,60,70,80,90,100,300,400,460])
plt.xlabel("Scores")
plt.ylabel("Count of Scores")
plt.title('Histogram of Altmetric Scores')
plt.show()
print ("Slope: ",slope, "Y - int: ",intercept, "R value: ",rvalue,"R^2: ", rva
print ("P value: ",pvalue, "Std error: ",stderr)
print ("R^2 Value is very low, thus linear regression is not a great choice")
print ("Optimal values (min. sum of squares) Altmetric Score: ",popt[0], "Stage
print ("Variance in Y - Values (Altmetric Scores): ",pcov[0,0])
print ("Covariance between Almetric Score and Stage of Cancer: ", pcov[1,0])
print ("Covariance score is postive, larger than R^2, thus indicating power reg
print ("Covariance Matrix Row 1", pcov[0])
print ("Covariance Matrix Row 2", pcov[1])
print ("Standard deviation (Sigma) of Altmetric Scores = ",perr[0])
```


Relationship between Altmetric Scores and Stage of Breast Cancer
450
400
100
80
35
30
25
20
15
10
5
0
Altmetric Score
1
2
3
4
Stage of Cancer
Linear Regression
Power Regression




```
Slope:  -6.723917672107885 Y - int:  25.28715400993614 R value:  -0.1265769956
5901473 R^2:  0.016021735830062233
P value:  0.17766251635443928 Std error:  4.957025291233932
R^2 Value is very low, thus linear regression is not a great choice
Optimal values (min. sum of squares) Altmetric Score:  21.134262589266086 Stag
e of Cancer:  1.9247412788518965
Variance in Y - Values (Altmetric Scores):  50.584731855121774
Covariance between Almetric Score and Stage of Cancer:  3.106389970302665
Covariance score is postive, larger than R^2, thus indicating power regression
is a good fit
Covariance Matrix Row 1 [50.58473186  3.10638997]
Covariance Matrix Row 2 [3.10638997 5.12180728]
Standard deviation (Sigma) of Altmetric Scores =  7.112294415666563
```

In [ ]: