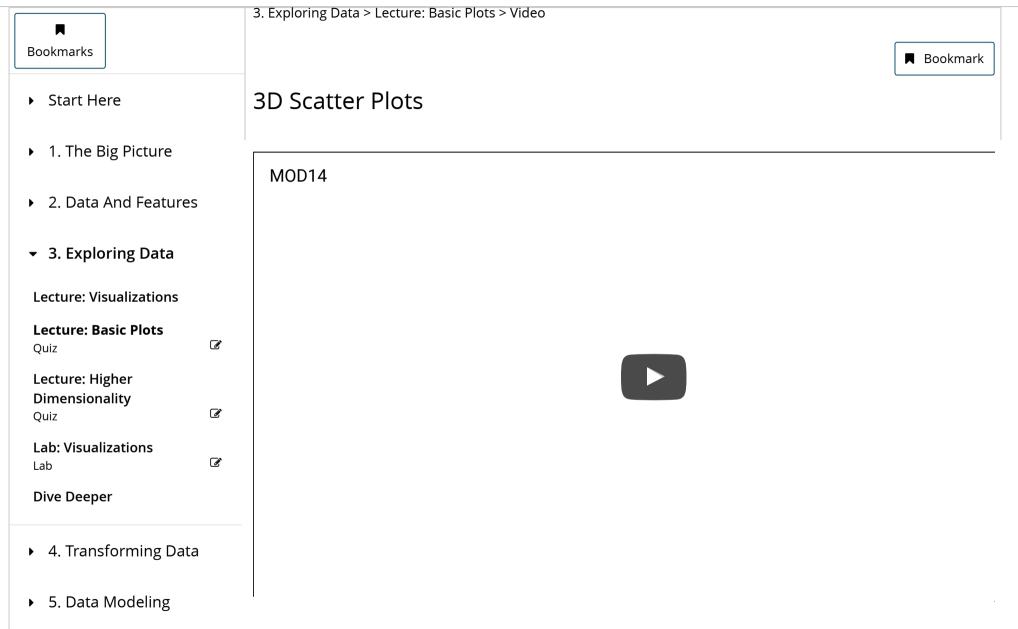


Microsoft: DAT210x Programming with Python for Data Science

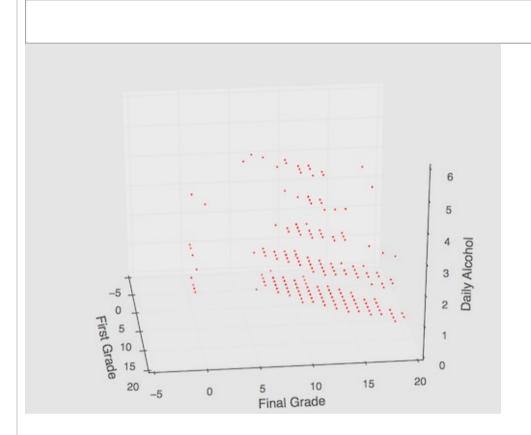


▶ 0:00 / 5:14

Download video
Download transcript
.srt

To follow up from the last section, there surely is a way to visualize the relationship between three variables simultaneously. That way is through 3D scatter plots. Unfortunately, the Pyplot member of Pandas dataframes don't natively support the ability to generate 3D plots... so for the sake of your visualization repertoire, you're going to learn how to make them directly with MatPlotLib.

```
import matplotlib
import matplotlib.pyplot as plt
from mpl toolkits.mplot3d import Axes3D
import pandas as pd
matplotlib.style.use('ggplot') # Look Pretty
# If the above line throws an error, use plt.style.use('gaplot') instead
student dataset = pd.read csv("/Datasets/students.data", index col=0)
fig = plt.figure()
ax = fig.add subplot(111, projection='3d')
ax.set xlabel('Final Grade')
ax.set ylabel('First Grade')
ax.set zlabel('Daily Alcohol')
ax.scatter(student_dataset.G1, student_dataset.G3, student_dataset['Dalc'], c='r',
marker='.')
plt.show()
```



This plot communicates a few things to you:

- It's still easy to see the positive, linear relationship between the first grade and the final grade. Generally the higher the first grade, the higher the final grade.
- A large number of the students in this study contume alcohol on a daily basis.
- The more alcohol a student consumes daily, the *worse* their final score is on average.
- Daily alcohol consumption doesn't really seem to be the feature that contributed to those select students who bombed their finals, scoring 0.

© All Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

















