Programming and Data Science for User Experience Researchers

HCDE 598 A — Jonathan T. Morgan & Ray Hong

Lecture | Week 2 | April 3, 2016

Group 1	Allie Deford	Morgan Duffy	Khuyen "Max" Lam	Daren Chaisy
Group 2	Ender Barillas	Jonathan Wang	Olivia Frederick	Yuan Zhuang
Group 3	Sahil Anand	Nick Throm	TJ Koines	George Karalis
Group 4	Nikhil Venkatesh	susan evans	Spencer James	Abhishek Anand
Group 5	Gabriela Madrid	Nick Anderson	John Jessup	Demi Boe
Group 6	David wang	Kristine Kohlhepp	Setumadhava G Kathawate	Emma Bulajewski
Group 7	Vishwas Shetty	Todd Little	Nicole Bernardi	Anne Burker

Administrativa

Practice quizzes

Re-scheduling Week 10

NEW TIME: Friday, June 3, 6pm

- Final presentations + course wrap-up & next steps
- Jonathan will provide pizza. Any volunteers for coffee/soda?

Everyone must present. There are three options:

1. *Come to Sieg Hall*

2. *Present remotely*

    i. *pizza/coffee not provided, and you have to stick around to support your classmates*

3. *NOT RECOMMENDED. Record your presentation ahead of time.*

    i. *No pizza, **and** you have to give a slightly longer presentation, **and** turn it in sooner, **and** you have to watch the other students presentations and give feedback on the discussion forum within 48 hours.*

*You **must** contact Jonathan by Friday, May 20 if you want to chose option 3.*

Your feedback

Week 1 feedback survey

- Overall pace
- Structured breaks
- Examples of final projects
- Peer programming/project time
- Tips for learning concepts, rules, and syntax
- Make learning to program more fun

Assignments & Grading

Participation

Participation

- 10% of your grade
- Based on engagement, not attendance

To assure that your absence does not negatively impact your grade:

1. *Review all materials we covered in the class*

2. *Perform any in-class exercises*

3. *Turn in any assignments due (including graded coding challenges)*

*You **must** contact Jonathan or Ray before 10pm the day after the class you missed.*

Tip: If you want to avoid falling behind, set up 30 minutes to talk with Jonathan or Ray.

http://wiki.communitydata.cc/DS4UX_(Spring_2016)#Grades
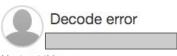
Graded Coding Challenges

Graded Coding Challenges

- Due: Week 5, Week 6, Week 7 —*by class time* (late work not accepted)
- 10% of your grade (3.3% each)

You have to attempt all questions to get full credit. You won't be graded on whether you're right or not.

http://wiki.communitydata.cc/DS4UX_(Spring_2016)#Final_Project_Idea

Tip: use the discussion board

Decode error

I just got this error:

```
UnicodeDecodeError: 'utf-8' codec can't decode bytes in position 7644-7645: invalid continuation byte
```

when trying to pull data from the csv data set I created. I also tried as tsv and got the same error. Google hasn't been much help in offering tips I can understand.

Here's the code I'm using:

```python
# -*- coding: utf8 -*-

from csv import DictReader

input_file = open("cascade_friends.tsv", "r")
import encoding_fix
import tweepy
from twitter_authentication import CONSUMER_KEY, CONSUMER_SECRET, ACCESS_TOKEN, ACCESS_TOKEN_SECRET

auth = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
auth.set_access_token(ACCESS_TOKEN, ACCESS_TOKEN_SECRET)

api = tweepy.API(auth)
output_file = open("user-bikes.csv","w")

#I have the users already, now I want to narrow down the list
tweetcount = 0
```

Tip: use the discussion board

Finding longest names

`1` `1`

I came up with this script for finding the longest names in the sets. I didn't find a single longest name, but I also had to do some hard coding of numbers to come up with that longest number. Here it is:

```
import ssadata

girls_names = ssadata.girls.keys()
boys_names = ssadata.boys.keys()
counter = 0

for name in girls_names:
    if len(name) >= 14: #how do I come up with a number that isn't hard coding it?
        print("the longest name is " + name)
        counter = counter + 1
        if counter >= len(name):
            print("\n")
print(counter)
for name in boys_names:
    if len(name) >= 14:
        print("the longest name is " + name)
counter = counter + 1

print(counter)

#Cycle through boys, return number

#Cycle through girls, return number
```

Final Project

Final project idea

- Due April 25th (week 5)
- ~600 words
- 10% of your grade

"Identify an area of interest, at least 2 datasets with relevant data, and at least 3-4 questions that you plan to explore."

http://wiki.communitydata.cc/DS4UX_(Spring_2016)#Final_Project_Idea

Final project proposal

- Due May 9 (week 7)
- ~1500 words
- 10% of your grade

"This proposal should focus on two questions: Why are you planning to do this analysis? What is your plan?"

http://wiki.communitydata.cc/DS4UX_(Spring_2016)#Final_Project_Proposal

Final project presentation

- Due June 3 (week 10 - CLASS ON FRIDAY)
- 5 minutes
- 10% of your grade

"Give a compelling talk and give me a sense for your project. Provide a very clear idea of what to expect in your final report."

http://wiki.communitydata.cc/DS4UX_(Spring_2016)#Final_Project_Presentation

Final project report

- Due June 8 (week 11 - Wednesday at 9pm)
- 3000 words
- 30% of your grade

build on the first two assignments & include detailed information on:

- *The problem or area you have identified*
- *Your research question(s)*
- *The methods, data, and approach that you used to collect the data*
- *The results and findings*
- *A discussion of limitations*
- *All your code*

http://wiki.communitydata.cc/DS4UX_(Spring_2016)#Final_Project_Report

Examples

https://nyssadatascience.wordpress.com/

Plagiarism

http://wiki.communitydata.cc/DS4UX_(Spring_2016)#Plagiarism

Questions?

Podium

Allie Deford
Morgan Duffy
Khuyen "Max" Lam
Daren Chaisy

Ender Barillas
Jonathan Wang
Olivia Frederick
Yuan Zhuang

Sahil Anand
Nick Throm
TJ Koines
George Karalis

Nikhil Venkatesh
Susan evans
Spencer James
Abhishek Anand

Gabriela Madrid
Nick Anderson
John Jessup
Demi Boe

David wang
Kristine Kohlhepp
Setumadhava G Kathawate
Emma Bulajewski

Vishwas Shetty
Todd Little
Nicole Bernardi
Anne Burker

+2 students