Lab 4 Submission

=========== Data Wrangler Script: CMSC

from wrangler import dw

import sys

if(len(sys.argv) < 3):

sys.exit('Error: Please include an input and output file. Example python script.py input.csv output.csv')

w = dw.DataWrangler()

# Split data repeatedly on newline into rows

w.add(dw.Split(column=["data"],

table=0,

status="active",

drop=True,

result="row",

update=False,

insert\_position="right",

row=None,

on="\n",

before=None,

after=None,

ignore\_between=None,

which=1,

max=0,

positions=None,

quote\_character=None))

# Wrap rows where data starts with 'CMSC'

w.add(dw.Wrap(column=[],

table=0,

status="active",

drop=False,

row=dw.Row(column=[],

table=0,

status="active",

drop=False,

conditions=[dw.StartsWith(column=[],

table=0,

status="active",

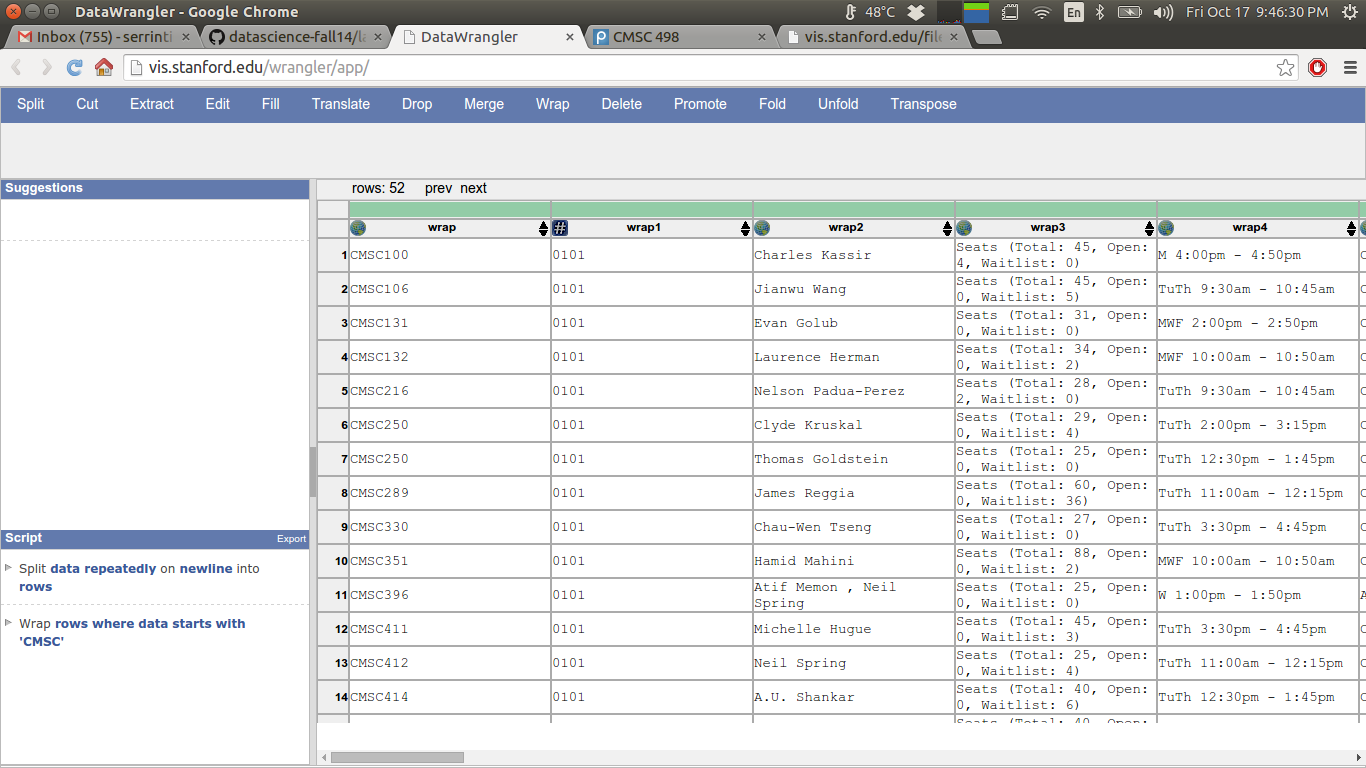
drop=False,

lcol="data",

value="CMSC",

op\_str="starts with")])))

w.apply\_to\_file(sys.argv[1]).print\_csv(sys.argv[2])



=========== Data Wrangler Script: World Cup 1

from wrangler import dw

import sys

if(len(sys.argv) < 3):

sys.exit('Error: Please include an input and output file. Example python script.py input.csv output.csv')

w = dw.DataWrangler()

# Split data repeatedly on newline into rows

w.add(dw.Split(column=["data"],

table=0,

status="active",

drop=True,

result="row",

update=False,

insert\_position="right",

row=None,

on="\n",

before=None,

after=None,

ignore\_between=None,

which=1,

max=0,

positions=None,

quote\_character=None))

# Wrap rows where data = '|-'

w.add(dw.Wrap(column=[],

table=0,

status="active",

drop=False,

row=dw.Row(column=[],

table=0,

status="active",

drop=False,

conditions=[dw.Eq(column=[],

table=0,

status="active",

drop=False,

lcol="data",

value="|-",

op\_str="=")])))

# Drop wrap

w.add(dw.Drop(column=["wrap"],

table=0,

status="active",

drop=True))

# Delete row 1

w.add(dw.Filter(column=[],

table=0,

status="active",

drop=False,

row=dw.Row(column=[],

table=0,

status="active",

drop=False,

conditions=[dw.RowIndex(column=[],

table=0,

status="active",

drop=False,

indices=[0])])))

# Cut from wrap1 before ' any word }'

w.add(dw.Cut(column=["wrap1"],

table=0,

status="active",

drop=False,

result="column",

update=True,

insert\_position="right",

row=None,

on=".\*",

before="[a-zA-Z]+}",

after=None,

ignore\_between=None,

which=1,

max=1,

positions=None))

# Cut from wrap1 after ' any uppercase word '

w.add(dw.Cut(column=["wrap1"],

table=0,

status="active",

drop=False,

result="column",

update=True,

insert\_position="right",

row=None,

on=".\*",

before=None,

after="[A-Z]+",

ignore\_between=None,

which=1,

max=1,

positions=None))

# Extract from wrap2 between '[\[' and ' FIFA'

w.add(dw.Extract(column=["wrap2"],

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on=".\*",

before=" FIFA",

after="\\[\\[",

ignore\_between=None,

which=1,

max=1,

positions=None))

# Extract from wrap2 between ' [\[' and ' FIFA'

w.add(dw.Extract(column=["wrap2"],

table=0,

status="active",

drop=False,

result="column",

update=False,

insert\_position="right",

row=None,

on=".\*",

before=" FIFA",

after=" \\[\\[",

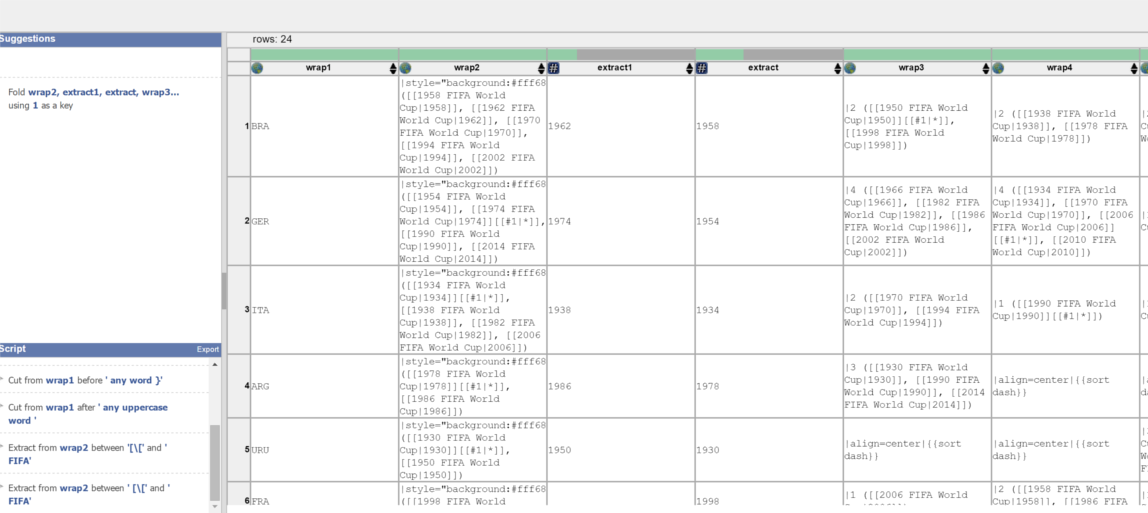
ignore\_between=None,

which=1,

max=1,

positions=None))

w.apply\_to\_file(sys.argv[1]).print\_csv(sys.argv[2])

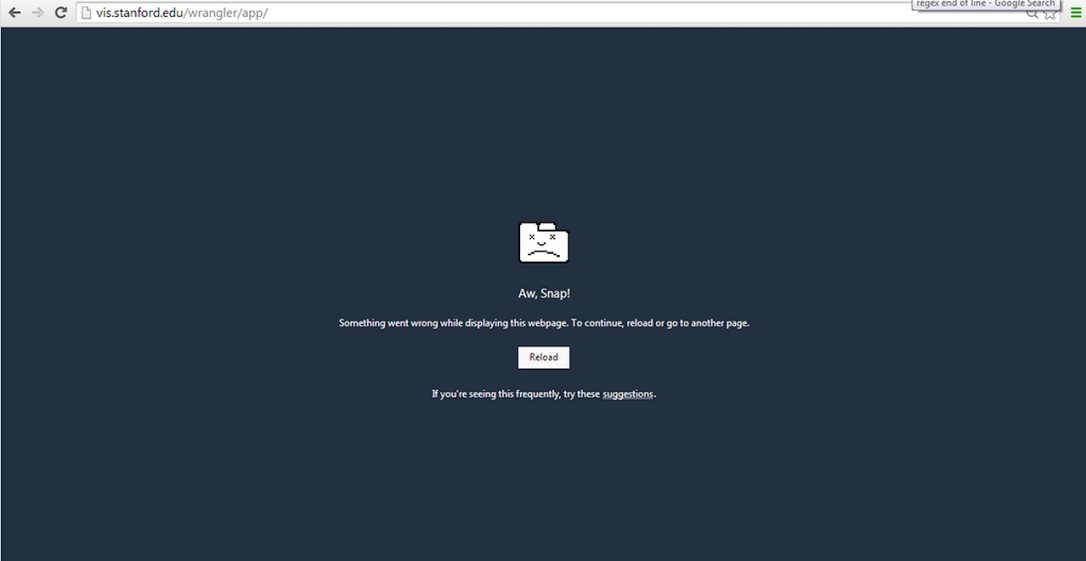


=========== Data Wrangler Script: World Cup 2

Data wrangler keeps crashing before I can finish World Cup 1.

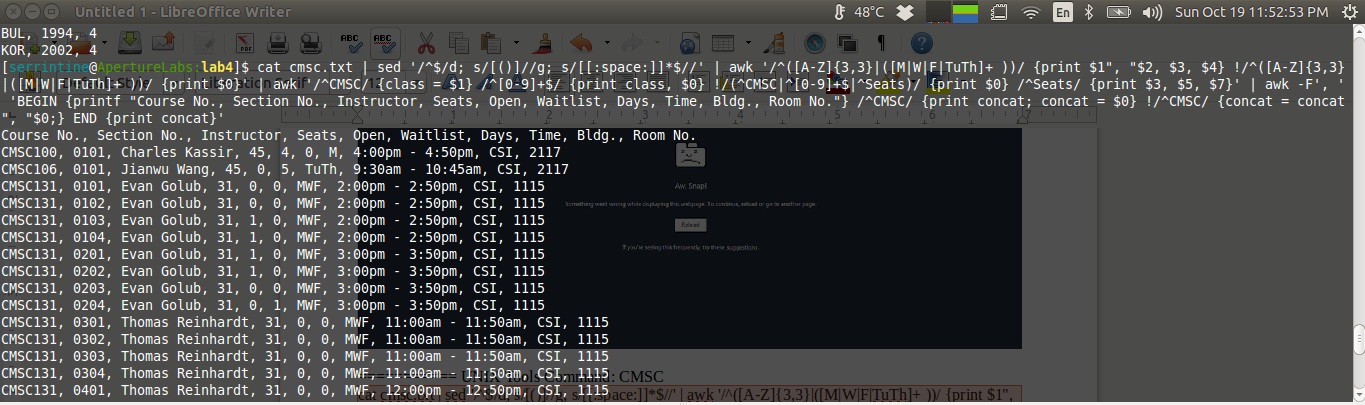
I have given up on moving forward with it. This is not even

remotely close to a competent tool for cleaning up data.



=========== UNIX Tools Command: CMSC

cat cmsc.txt | sed '/^$/d; s/[()]//g; s/[[:space:]]\*$//' | awk '/^([A-Z]{3,3}|([M|W|F|TuTh]+ ))/ {print $1", "$2, $3, $4} !/^([A-Z]{3,3}|([M|W|F|TuTh]+ ))/ {print $0}' | awk '/^CMSC/ {class = $1} /^[0-9]+$/ {print class, $0} !/(^CMSC|^[0-9]+$|^Seats)/ {print $0} /^Seats/ {print $3, $5, $7}' | awk -F', ' 'BEGIN {printf "Course No., Section No., Instructor, Seats, Open, Waitlist, Days, Time, Bldg., Room No."} /^CMSC/ {print concat; concat = $0} !/^CMSC/ {concat = concat", "$0;} END {print concat}'



=========== UNIX Tools Command: World Cup 1

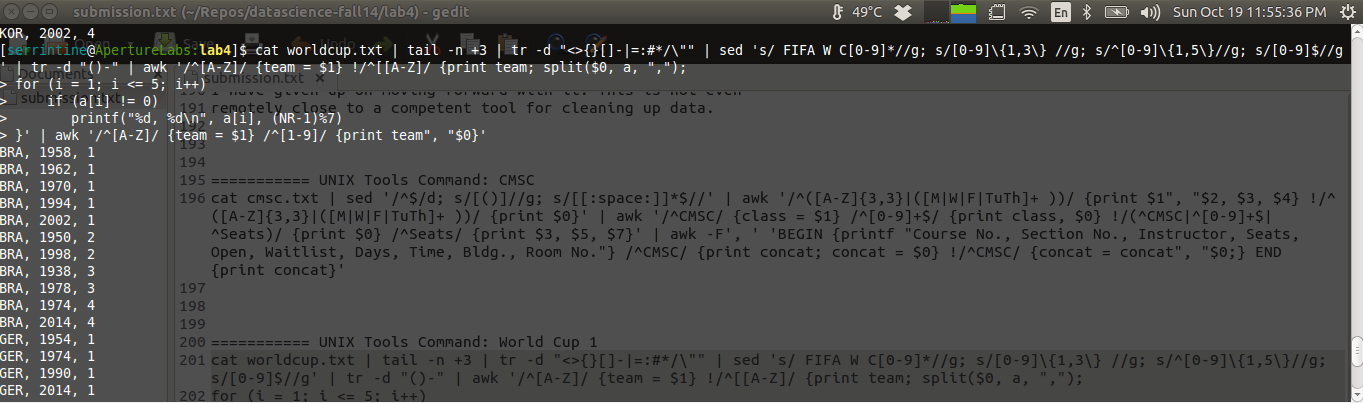
cat worldcup.txt | tail -n +3 | tr -d "<>{}[]-|=:#\*/\"" | sed 's/ FIFA W C[0-9]\*//g; s/[0-9]\{1,3\} //g; s/^[0-9]\{1,5\}//g; s/[0-9]$//g' | tr -d "()-" | awk '/^[A-Z]/ {team = $1} !/^[[A-Z]/ {print team; split($0, a, ",");

for (i = 1; i <= 5; i++)

if (a[i] != 0)

printf("%d, %d\n", a[i], (NR-1)%7)

}' | awk '/^[A-Z]/ {team = $1} /^[1-9]/ {print team", "$0}'



=========== Python Script: CMSC

=========== Python Script: World Cup 1

=========== Python Script: World Cup 2