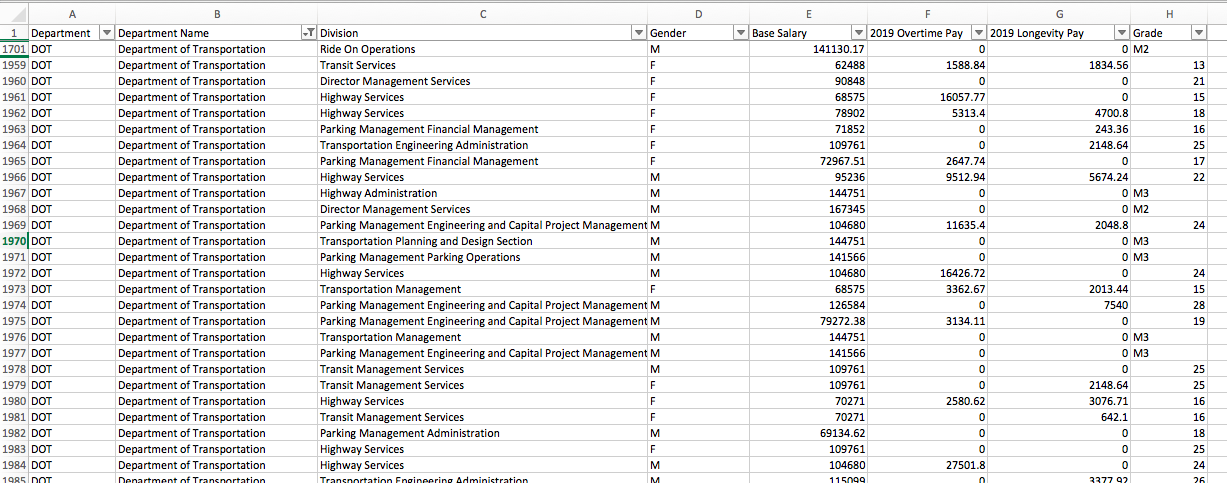
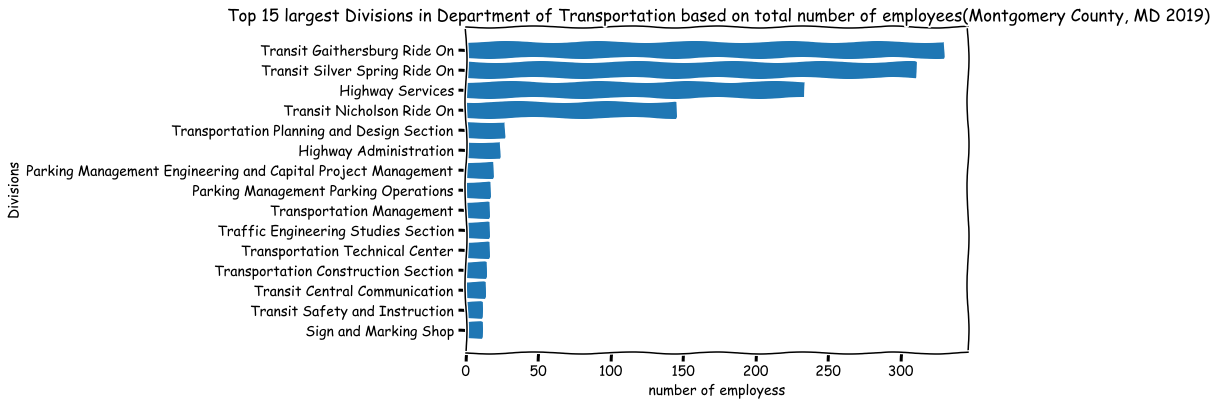
Name : JUNYI LI

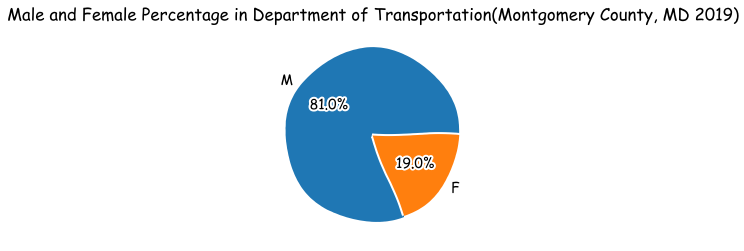
Class: CISC 504-50- A-2020

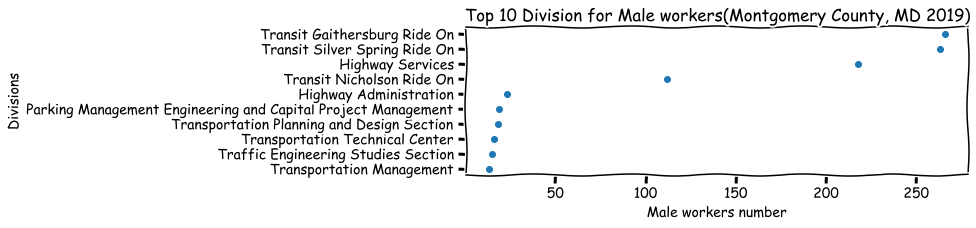
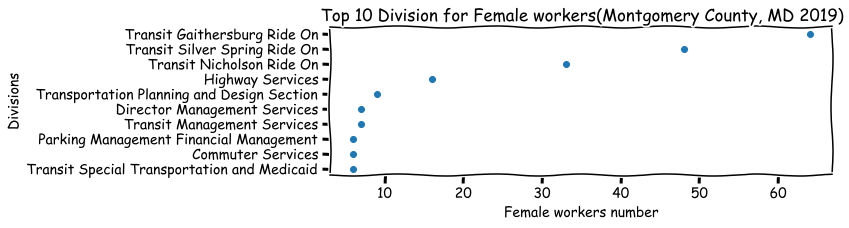
Date: 01/22/2021

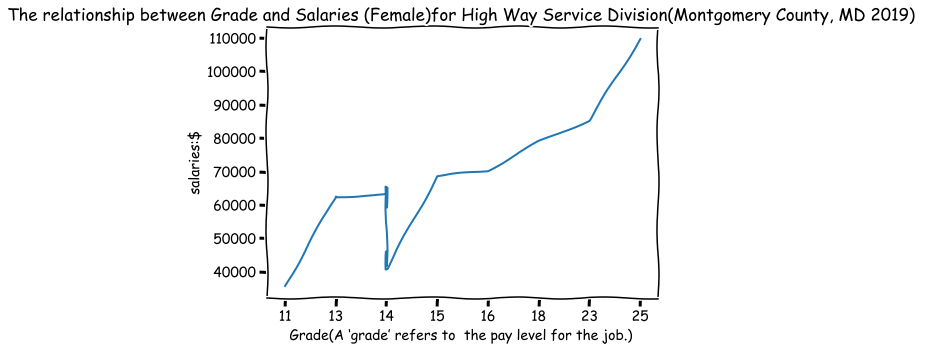
Link : https://catalog.data.gov/dataset/employee-salaries-2019

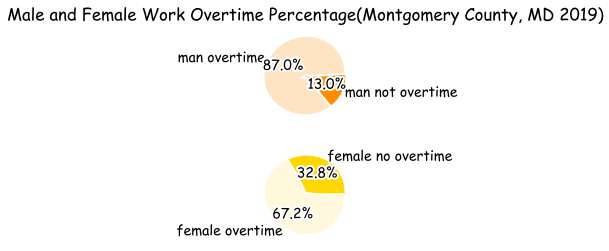


Graph:









Code:

# Create an new csv document with only dataset from Department of Transportation by Pandas

import csv

import numpy as np

from collections import Counter

from matplotlib import pyplot as plt

import pandas as pd

T\_df=pd.read\_csv('Employee\_Salaries\_-\_2019.csv')

filt=T\_df['Department'].str.contains('DOT',na=False)

w\_df=T\_df.loc[filt] #filter the Department that is transportation

w\_df.to\_csv('transportation.csv', index=False)

# 1. Top 5 largest Divisions in Department of Transportation based on number of employees.

plt.xkcd()

with open('transportation.csv') as csv\_file:

csv\_reader=csv.DictReader(csv\_file)#read csv

dep\_counter=Counter()

for row in csv\_reader:

dep\_counter.update(row['Division'].split(';'))

Department=[]

num\_people=[]

for item in dep\_counter.most\_common(15):#top 5 departments

Department.append(item[0])

num\_people.append(item[1])

Department.reverse() # from largest to lowest

num\_people.reverse()

plt.figure(figsize=(9,6))

plt.barh(Department,num\_people)# add h to bar

plt.title('Top 15 largest Divisions in Department of Transportation based on total number of employees(Montgomery County, MD 2019)')

plt.ylabel('Divisions')

plt.xlabel('number of employess')

plt.savefig('all divsion',bbox\_inches = 'tight')

# 2. Male and Female Percentages in Department of Transportation.

plt.xkcd()

with open('transportation.csv') as csv\_file:

csv\_reader=csv.DictReader(csv\_file)#read csv

gen\_counter=Counter()

for row in csv\_reader:

gen\_counter.update(row['Gender'].split(';'))#the way to divide them

gender=[]

num\_people=[]

for item in gen\_counter.most\_common(5):

gender.append(item[0])

num\_people.append(item[1])

plt.pie(num\_people, labels=gender,autopct='%1.1f%%')

plt.title('Male and Female Percentage in Department of Transportation(Montgomery County, MD 2019)')

plt.savefig('workers percentage',bbox\_inches = 'tight')

3. # Create Female workers only csv

plt.xkcd()

df=pd.read\_csv('transportation.csv')

f\_df=df[df['Gender']=='F'] #filter out the female

f\_df.to\_csv('Female\_workers.csv', index=False)

with open('Female\_workers.csv') as csv\_file:

csv\_reader=csv.DictReader(csv\_file)#read csv

gen\_counter=Counter()

for row in csv\_reader:

gen\_counter.update(row['Division'].split(';'))

division=[]

num\_people=[]

for item in gen\_counter.most\_common(10):

division.append(item[0])

num\_people.append(item[1])

# Create Male workers only csv

plt.xkcd()

df=pd.read\_csv('transportation.csv')

f\_df=df[df['Gender']=='M']

f\_df.to\_csv('Male\_workers.csv', index=False)

with open('Male\_workers.csv') as csv\_file:

csv\_reader=csv.DictReader(csv\_file)#read csv

m\_counter=Counter()

for row in csv\_reader:

m\_counter.update(row['Division'].split(';'))

m\_division=[]

m\_num\_people=[]

for item in m\_counter.most\_common(10):

m\_division.append(item[0])

m\_num\_people.append(item[1])

m\_division.reverse()

m\_num\_people.reverse()

plt.figure(figsize=(9,6))

plt.subplot(2,1,1)

plt.scatter(m\_num\_people, m\_division)

plt.title('Top 10 Division for Male workers(Montgomery County, MD 2019)')

plt.ylabel('Divisions')

plt.xlabel('Male workers number')

plt.savefig('m\_d',bbox\_inches = 'tight')

division.reverse()

num\_people.reverse()

plt.figure(figsize=(9,6))

plt.subplot(2,1,2)

plt.scatter(num\_people,division)

plt.title('Top 10 Division for Female workers(Montgomery County, MD 2019)')

plt.ylabel('Divisions')

plt.xlabel('Female workers number')

plt.savefig('F\_d',bbox\_inches = 'tight')

print(m\_division,m\_num\_people)

#4. The relationship between Grade and Salaries (Female)for High Way Service Division

plt.xkcd()

f=pd.read\_csv('transportation.csv')

# print out, the leftest numbers are indices for the row

High=f[(f.Division=='Highway Services')& (f.Gender=='F')]

High\_new=High.rename(columns={'Base Salary':'Base\_Salary'}).sort\_values(by='Grade')

plt.figure(figsize=(7,5))

plt.plot(High\_new.Grade, High\_new.Base\_Salary)

plt.title('The relationship between Grade and Salaries (Female)for High Way Service Division(Montgomery County, MD 2019) ')

plt.xlabel('Grade(A ‘grade’ refers to the pay level for the job.) ')

plt.ylabel('salaries:$')

plt.savefig('grade', bbox\_inches = 'tight')

#5 male and female overtime percentage respectl

m=pd.read\_csv('Male\_workers.csv')

a\_m=m.rename(columns={'Base Salary':'Base\_Salary','2019 Overtime Pay':'2019\_Overtime\_Pay'})

all\_men=len(a\_m.Division) # total male workers

#print(all\_men)

om=a\_m[a\_m['2019\_Overtime\_Pay']>=1]

overtime\_man=len(om.Division)

#print(overtime\_man)

f=pd.read\_csv('Female\_workers.csv')

a\_f=f.rename(columns={'Base Salary':'Base\_Salary','2019 Overtime Pay':'2019\_Overtime\_Pay'})

all\_female=len(a\_f.Division) #total female workers

#print(all\_female)

fm=a\_f[a\_f['2019\_Overtime\_Pay']>=1]

overtime\_female=len(fm.Division)

#print(overtime\_female)

plt.savefig('grade')

slices=[overtime\_man,all\_men-overtime\_man]

labels=['man overtime','man not overtime']

colorsm=['bisque','darkorange']

plt.subplot(2,1,1)

plt.pie(slices, labels=labels,autopct='%1.1f%%',colors=colorsm)

plt.title('Male and Female Work Overtime Percentage(Montgomery County, MD 2019)')

#plt.savefig('overtime percentage1',bbox\_inches = 'tight')

f\_slices=[all\_female-overtime\_female,overtime\_female]

f\_labels=['female no overtime','female overtime']

colorsf=['gold','cornsilk']

plt.subplot(2,1,2)

plt.pie(f\_slices, labels=f\_labels,autopct='%1.1f%%',colors=colorsf)

plt.savefig('Male and Female overtime percentage2',bbox\_inches = 'tight')