

## Gabriel Krishnadasan

### Module 1 Assignment

1. The Exercise 1 Dataset (located in your assignment prompt in Canvas) contains a portion of the data from NYC about causes of death for the year 2010.

Dataset format: CSV

Field names (in order): Year, Ethnicity, Sex, Cause of Death, Death Count.

Answer the following questions from this data using UNIX commands.

- 1.1: How many male record groups and how many female record groups does the data have? (8 points)

```
gabekrishnadasan@gKrishnadasan Assignment1 % cut -d"," -f3 deaths.csv | sort | uniq -c
161 Female
163 Male
```

- 1.2: How many white female groups are there? Copy entire records of females to a new text file where the records are organized by death count in descending order. (8 points)

```
gabekrishnadasan@gKrishnadasan Assignment1 % grep "Female" deaths.csv > femaleDeaths.csv
gabekrishnadasan@gKrishnadasan Assignment1 % cut -d"," -f2 femaleDeaths.csv | sort | uniq -c | sort -nr
36 White
36 Hispanic
35 Black
32 Asian
21 Unknown
1 Other
```

1.3: List all causes of death by their counts in descending order; do not worry about summing any rows together, just sort the count of causes of death column. What are the three most common causes of death for black males, and five least common causes of death for hispanic females? (10 points)

```
gabekrishnadasan@gKrishnadasan Assignment1 % cut -d',' -f4 deaths.csv | sort | uniq -c | sort -nr
11 Accidents Except Drug Poisoning
10 Viral Hepatitis
10 Septicemia
10 Nephritis and Nephrotic Syndromes
10 Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco
10 Malignant Neoplasms (cancer)
10 Leukemia
10 Intentional Self-Harm (Suicide)
10 Influenza (Flu) and Pneumonia
10 Essential Hypertension and Renal Diseases
10 Diseases of Heart
10 Diabetes Mellitus
10 Chronic Lower Respiratory Diseases
10 Certain Conditions Originating in the Perinatal Period
10 Cerebrovascular Disease (stroke)
10 Atherosclerosis
9 Peptic Ulcer
9 Parkinson's Disease
9 Insitu or Benign / Uncertain Neoplasms
9 Human Immunodeficiency Virus (HIV) Disease
9 Chronic Liver Disease and Cirrhosis
9 Assault (Homicide)
9 Aortic Aneurysm and Dissection
9 Alzheimer's Disease
8 Tuberculosis (TB)
8 Non-Hodgkins's Lymphoma
8 Mental and Behavioral Disorders due to use of Alcohol
8 Congenital Malformations and Chromosomal Abnormalities
8 Complications of Medical and Surgical Care
8 Cholelithiasis and Disorders of Gallbladder
8 Anemias
8 All Censored Causes
7 Meningitis
6 Pneumonitis due to Solids and Liquids
5 Pregnancy and Childbirth
5 Maternal Causes
4 Legal Intervention
```

```
gabekrishnadasan@gKrishnadasan Assignment1 % grep "Black, Male" deaths.csv > blackMaleDeaths.csv
gabekrishnadasan@gKrishnadasan Assignment1 % cut -d',' -f4 blackMaleDeaths.csv | sort | uniq -c | sort -nr | head -3
1 Viral Hepatitis
1 Tuberculosis (TB)
1 Septicemia
```

```
gabekrishnadasan@gKrishnadasan Assignment1 % grep "Hispanic" femaleDeaths.csv | cut -d',' -f4 | sort | uniq -c | sort -nr | head -5
1 Viral Hepatitis
1 Tuberculosis (TB)
1 Septicemia
1 Pregnancy and Childbirth
1 Pneumonitis due to Solids and Liquids
```

2. Obtained from UNICEF, the Exercise 2 Dataset (located in your assignment prompt in Canvas) contains data related to the population of 70+ countries for the year 2017.

Dataset format: CSV

Field names (in order): Country, Population, Urban Population, Percentage of Urban Population.

Answer the following questions from this data using UNIX commands:

2.1: Which country has the lowest percentage of urban population? (8 points)

```
gabekrishnadasan@gKrishnadasan Assignment1 % cut -d',' -f1,4 populations.csv | sort -t',' -n -k2 | head -1  
Burundi,13
```

2.2: List the countries where the urban population is more than 10 million and yet they comprise less than half of the population. (10 points)

```
gabekrishnadasan@gKrishnadasan Assignment1 % awk -F',' '$3 >= 10000000 && $4 < 50' populations.csv | cut -d',' -f1  
Bangladesh  
Democratic Republic of the Congo  
Egypt  
Ethiopia  
India  
Kenya  
Myanmar  
Pakistan  
Philippines  
Sudan  
Thailand  
United Republic of Tanzania  
Viet Nam  
Yemen
```

3. For the following exercise, use the Exercise 3 Dataset (located in your assignment prompt in Canvas), which contains the availability of essential medicines in 38 countries for the years 2007 - 2013, obtained from the World Health Organization (WHO).

Dataset format: CSV

Field names (in order): Country, Median availability of selected generic medicines (%) - Private, Median availability of selected generic medicines (%) - Public

Answer the following questions from this data using UNIX commands:

3.1: Which country had the lowest percentage median availability of selected generic medicines in private. (8 points)

```
gabekrishnadasan@gKrishnadasan Assignment1 % tail -n+3 medicines.csv | cut -d'"'"' -f2,4,6 | sed 's"/,/g' > medicines2.txt  
gabekrishnadasan@gKrishnadasan Assignment1 % cut -d',' -f1,2 medicines2.txt | sort -t',' -k2 -n | head -1  
India,2.8
```

3.2: Top five countries with highest public percentage median availability of selected generic medicines. (8 points)

```
gabekrishnadasan@gKrishnadasan Assignment1 % tail -n+3 medicines.csv | cut -d'"'"' -f2,4,6 | sed 's"/,/g' > medicines2.txt  
gabekrishnadasan@gKrishnadasan Assignment1 % cut -d',' -f1,3 medicines2.txt | sort -t',' -k2 -nr | head -5  
Russian Federation,100.0  
Cook Islands,100.0  
Oman,96.7  
Iran (Islamic Republic of),96.7  
Syrian Arab Republic,93.0
```

3.3: List the top three countries where it is best to rely on the private availability of selected generic medicines than public. Explain your answer with valid reasons. (10 points)

```
gabekrishnadasan@gKrishnadasan Assignment1 % tail -n+3 medicines.csv | cut -d'"'"' -f2,4,6 | sed 's"/",/g' > medicines2.txt
gabekrishnadasan@gKrishnadasan Assignment1 % awk -F',' '{print $1, $2-$3}' medicines2.txt | sort -k2 -nr | head -3
Brazil 76.7
Kyrgyzstan 70
Mongolia 51.3
```

My reason for these three countries is that the difference in the private and the public is the greatest, and it is best to rely on the private availability of medicines.

4. Write a Python script that assigns the list [25,18,9,13,34,15,22,17,12,37,15] to a variable "age" and uses that information to:

4.1: Determine if a person is in high school or not. Assume that for a person to be in high school, their age should be between 14 and 18, inclusive. (5 points)

4.2: From the list, calculate the percentage of people not going to high school. (5 points)

```
1 age = [25,18,9,13,34,15,22,17,12,37,15]
2 notInHSCount = 0
3
4 for i in range(len(age)):
5     if age[i] >= 14 and age[i] <= 18:
6         print("Index ", i, " is in High School")
7     else:
8         print("Index ", i, " is not in High School")
9         notInHSCount += 1
10
11 print("The percentage of people not in High School is: ", notInHSCount / len(age))
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL JUPYTER

```
/usr/bin/python3 /Users/gabekrishnadasan/Desktop/2024S/COMP352/Material/Assignment1/age.py
gabekrishnadasan@gKrishnadasan Material % /usr/bin/python3 /Users/gabekrishnadasan/Desktop/2024S/COMP352/Material/Assignment1/age.py
Index 0 is not in High School
Index 1 is in High School
Index 2 is not in High School
Index 3 is not in High School
Index 4 is not in High School
Index 5 is in High School
Index 6 is not in High School
Index 7 is in High School
Index 8 is not in High School
Index 9 is not in High School
Index 10 is in High School
The percentage of people not in High School is: 0.6363636363636364
```

