

Module Two Class Assignment

Mayel Espino

Department of Data Science University of San Diego

ADS-500B-01-SP21 - Data Science Programming

M.S. Jules Malin

March 11, 2021

Module Two Class Assignment

Exercises

Exercise One

Dataset (located in your assignment prompt in Blackboard) contains apportion 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?

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.

1.3: List all causes of death by their frequencies in descending order. What are the three most frequent causes of death for black males, and five least frequent causes of death for Hispanic females?

Answer 1.1

If grouped by Ethnicity there are 6 Male groups, and their counts are the second column

```
~/REPOS/usd-ads/ads500b/module02 >>> grep Male Exercise1_data.csv | tr ' ' '_' | awk -F, '{arr[$2]++}END{for (a in arr) print a, arr[a]}'
Asian 34
Other 1
Black 35
Hispanic 35
White 35
Unknown 23
~/REPOS/usd-ads/ads500b/module02 >>>
```

If grouped by cause of death there are 35 Male groups, and their counts are the second column:

```
~/REPOS/usd-ads/ads500b/module02 >>> grep Male Exercise1_data.csv |tr ' ' '_' | awk -F, '{arr[$4]++}END{for (a in arr) print a, arr[a]}'
Human_Immunodeficiency_Virus_(HIV)_Disease 5
Legal_Intervention 4
Peptic_Ulcer 5
Insitu_or_Benign/_Uncertain_Neoplasms 4
Leukemia 5
Septicemia 5
Cerebrovascular_Disease_(stroke) 5
Non-Hodgkins's_Lymphoma 4
Chronic_Liver_Disease_and_Cirrhosis 5
Diabetes_Mellitus 5
Intentional_Self-Harm_(Suicide) 5
Atherosclerosis 5
Meningitis 4
Assault_(Homicide) 5
Influenza_(Flu)_and_Pneumonia 5
All_Censored_Causes 4
Diseases_of_Heart 5
Essential_Hypertension_and_Renal_Diseases 5
Tuberculosis_(TB) 4
Malignant_Neoplasms_(cancer) 5
Mental_and_Behavioral_Disorders_due_to_use_of_or_Accidental_Poisoning_by_Psychoactive_Substances_Excluding_Alcohol_and_Tobacco 5
Cholelithiasis_and_Disorders_of_Gallbladder 4
Mental_and_Behavioral_Disorders_due_to_use_of_Alcohol 5
Complications_of_Medical_and_Surgical_Care 5
Alzheimer's_Disease 4
Parkinson's_Disease 4
Certain_Conditions_Originating_in_the_Perinatal_Period 5
Viral_Hepatitis 5
Accidents_Except_Drug_Poisoning 6
Nephritis_and_Nephrotic_Syndromes 5
Pneumonitis_due_to_Solids_and_Liquids 3
Congenital_Malformations_and_Chromosomal_Abnormalities 4
Anemias 4
Aortic_Aneurysm_and_Dissection 5
Chronic_Lower_Respiratory_Diseases 5
~/REPOS/usd-ads/ads500b/module02 >>>
```

If grouped by ethnicity, there are 6 Female groups, and their counts are the second column:

```
~/REPOS/usd-ads/ads500b/module02 >>> grep Female Exercise1_data.csv |tr ' ' '_' | awk -F, '{arr[$2]++}END{for (a in arr) print a, arr[a]}'
Asian 32
Other 1
Black 35
Hispanic 36
White 36
Unknown 21
~/REPOS/usd-ads/ads500b/module02 >>>
```

If grouped by cause of death there are 36 Female groups, and their counts are the second column:

```
~/REPOS/usd-ads/ads500b/module02 >>> grep Female Exercise1_data.csv |tr ' ' '_' | awk -F, '{arr[$4]++}END{for (a in arr) print a, arr[a]}'
Human_Immunodeficiency_Virus_(HIV)_Disease 4
Peptic_Ulcer 4
Insitu_or_Benign/_Uncertain_Neoplasms 5
Leukemia 5
Septicemia 5
Cerebrovascular_Disease_(stroke) 5
Non-Hodgkins's_Lymphoma 4
Pregnancy_and_Childbirth 5
Chronic_Liver_Disease_and_Cirrhosis 4
Diabetes_Mellitus 5
Intentional_Self-Harm_(Suicide) 5
Atherosclerosis 5
Meningitis 3
Assault_(Homicide) 4
Influenza_(Flu)_and_Pneumonia 5
All_Censored_Causes 4
Diseases_of_Heart 5
Essential_Hypertension_and_Renal_Diseases 5
Tuberculosis_(TB) 4
Malignant_Neoplasms_(cancer) 5
Mental_and_Behavioral_Disorders_due_to_use_of_or_Accidental_Poisoning_by_Psychoactive_Substances_Excluding_Alcohol_and_Tobacco 5
Cholelithiasis_and_Disorders_of_Gallbladder 4
Mental_and_Behavioral_Disorders_due_to_use_of_Alcohol 3
Complications_of_Medical_and_Surgical_Care 3
Alzheimer's_Disease 5
Parkinson's_Disease 5
Certain_Conditions_Originating_in_the_Perinatal_Period 5
Viral_Hepatitis 5
Accidents_Except_Drug_Poisoning 5
Nephritis_and_Nephrotic_Syndromes 5
Pneumonitis_due_to_Solids_and_Liquids 3
Congenital_Malformations_and_Chromosomal_Abnormalities 4
Anemias 4
Maternal_Causes 5
Aortic_Aneurysm_and_Dissection 4
Chronic_Lower_Respiratory_Diseases 5
~/REPOS/usd-ads/ads500b/module02 >>>
```

Answer 1.2

There are 36 groups, when grouped by cause of death, for white Females:

```
~/REPOS/usd-ads/ads500b/module02 >>> grep Female Exercise1_data.csv | grep White | tr ' ' '_' > white_females.txt
~/REPOS/usd-ads/ads500b/module02 >>> grep Female Exercise1_data.csv | grep White | tr ' ' '_' | awk -F, '{arr[$4]++}END{for (a in arr) print a, arr[a]}' | wc -l
36
```

They are:

```
~/REPOS/usd-ads/ads500b/module02 >>> grep Female Exercise1_data.csv | grep White | tr ' ' '_' | awk -F, '{arr[$4]++}END{for (a in arr) print a, arr[a]}'
Human_Immunodeficiency_Virus_(HIV)_Disease 1
Peptic_Ulcer 1
Insitu_or_Benign/_/_Uncertain_Neoplasms 1
Leukemia 1
Septicemia 1
Cerebrovascular_Disease_(stroke) 1
Non-Hodgkins's_Lymphoma 1
Pregnancy_and_Childbirth 1
Chronic_Liver_Disease_and_Cirrhosis 1
Diabetes_Mellitus 1
Intentional_Self-Harm_(Suicide) 1
Atherosclerosis 1
Meningitis 1
Assault_(Homicide) 1
Influenza_(Flu)_and_Pneumonia 1
All_Censored_Causes 1
Diseases_of_Heart 1
Essential_Hypertension_and_Renal_Diseases 1
Tuberculosis_(TB) 1
Malignant_Neoplasms_(cancer) 1
Mental_and_Behavioral_Disorders_due_to_use_of_or_Accidental_Poisoning_by_Psychoactive_Substances_Excluding_Alcohol_and_Tobacco 1
Cholelithiasis_and_Disorders_of_Gallbladder 1
Mental_and_Behavioral_Disorders_due_to_use_of_Alcohol 1
Complications_of_Medical_and_Surgical_Care 1
Alzheimer's_Disease 1
Certain_Conditions_Originating_in_the_Perinatal_Period 1
Parkinson's_Disease 1
Viral_Hepatitis 1
Accidents_Except_Drug_Poisoning 1
Nephritis_and_Nephrotic_Syndromes 1
Pneumonitis_due_to_Solids_and_Liquids 1
Congenital_Malformations_and_Chromosomal_Abnormalities 1
Anemias 1
Maternal_Causes 1
Aortic_Aneurysm_and_Dissection 1
Chronic_Lower_Respiratory_Diseases 1
~/REPOS/usd-ads/ads500b/module02 >>>
```

Copy entire records for white females to a text file, organized by death count:

```
~/REPOS/usd-ads/ads500b/module02 >>> grep Female Exercise1_data.csv | grep White | sort -t, -r -n -k5 > white_females.txt
~/REPOS/usd-ads/ads500b/module02 >>> cat white_females.txt
2010,White,Female,Diseases of Heart,5351
2010,White,Female,Malignant Neoplasms (cancer),3438
2010,White,Female,Influenza (Flu) and Pneumonia,707
2010,White,Female,Chronic Lower Respiratory Diseases,501
2010,White,Female,Cerebrovascular Disease (stroke),382
2010,White,Female,Diabetes Mellitus,258
2010,White,Female,Alzheimer's Disease,247
2010,White,Female,Essential Hypertension and Renal Diseases,219
2010,White,Female,Accidents Except Drug Poisoning,163
2010,White,Female,Leukemia,139
2010,White,Female,Non-Hodgkins's Lymphoma,127
2010,White,Female,Septicemia,107
2010,White,Female,Nephritis and Nephrotic Syndromes,90
2010,White,Female,Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco,75
2010,White,Female,Atherosclerosis,74
2010,White,Female,Insitu or Benign / Uncertain Neoplasms,69
2010,White,Female,Chronic Liver Disease and Cirrhosis,67
2010,White,Female,Intentional Self-Harm (Suicide),61
2010,White,Female,Parkinson's Disease,55
2010,White,Female,Viral Hepatitis,32
2010,White,Female,Congenital Malformations and Chromosomal Abnormalities,32
2010,White,Female,Certain Conditions Originating in the Perinatal Period,31
2010,White,Female,Human Immunodeficiency Virus (HIV) Disease,24
2010,White,Female,Peptic Ulcer,22
2010,White,Female,Aortic Aneurysm and Dissection,22
2010,White,Female,Cholelithiasis and Disorders of Gallbladder,19
2010,White,Female,Mental and Behavioral Disorders due to use of Alcohol,16
2010,White,Female,Assault (Homicide),13
2010,White,Female,Anemias,6
2010,White,Female,Pneumonitis due to Solids and Liquids,5
2010,White,Female,Complications of Medical and Surgical Care,5
2010,White,Female,Meningitis,3
2010,White,Female,Pregnancy and Childbirth,2
2010,White,Female,Maternal Causes,2
2010,White,Female,All Censored Causes,2
2010,White,Female,Tuberculosis (TB),1
~/REPOS/usd-ads/ads500b/module02 >>>
```

Answer 1.3

List of all the causes of death in descending order of frequency

```
~/REPOS/usd-ads/ads500b/module02 >>> cat Exercise1_data.csv | sort -t, -r -n -k5,5 -k2,2 -k3,3
2010,White,Female,Diseases of Heart,5351
2010,White,Male,Diseases of Heart,4495
2010,White,Female,Malignant Neoplasms (cancer),3438
2010,White,Male,Malignant Neoplasms (cancer),3408
2010,Black,Female,Diseases of Heart,2282
2010,Black,Male,Diseases of Heart,2015
2010,Black,Female,Malignant Neoplasms (cancer),1768
2010,Black,Male,Malignant Neoplasms (cancer),1540
2010,Hispanic,Male,Diseases of Heart,1354
2010,Hispanic,Female,Diseases of Heart,1316
2010,Hispanic,Male,Malignant Neoplasms (cancer),1057
2010,Hispanic,Female,Malignant Neoplasms (cancer),1045
2010,White,Female,Influenza (Flu) and Pneumonia,707
2010,White,Male,Influenza (Flu) and Pneumonia,614
2010,Asian,Male,Malignant Neoplasms (cancer),526
2010,White,Female,Chronic Lower Respiratory Diseases,501
2010,Asian,Male,Diseases of Heart,479
2010,Asian,Female,Malignant Neoplasms (cancer),415
2010,Asian,Female,Diseases of Heart,394
2010,White,Male,Chronic Lower Respiratory Diseases,388
2010,Black,Female,Diabetes Mellitus,384
2010,White,Female,Cerebrovascular Disease (stroke),382
2010,Black,Male,Assault (Homicide),299
2010,Black,Male,Human Immunodeficiency Virus (HIV) Disease,297
2010,Black,Male,Diabetes Mellitus,279
2010,White,Male,Cerebrovascular Disease (stroke),276
2010,Black,Female,Influenza (Flu) and Pneumonia,275
2010,White,Male,Accidents Except Drug Poisoning,271
2010,Black,Female,Cerebrovascular Disease (stroke),259
2010,White,Female,Diabetes Mellitus,258
2010,Black,Female,Essential Hypertension and Renal Diseases,252
2010,White,Female,Alzheimer's Disease,247
2010,White,Male,Diabetes Mellitus,245
2010,Hispanic,Male,Influenza (Flu) and Pneumonia,235
2010,Hispanic,Female,Influenza (Flu) and Pneumonia,228
2010,White,Female,Essential Hypertension and Renal Diseases,219
2010,Hispanic,Female,Diabetes Mellitus,219
2010,White,Male,Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco,211
2010,Black,Female,Chronic Lower Respiratory Diseases,206
2010,Black,Male,Chronic Lower Respiratory Diseases,202
2010,Black,Male,Influenza (Flu) and Pneumonia,201
2010,White,Male,Leukemia,194
```


2010, Hispanic, Male, Human Immunodeficiency Virus (HIV) Disease, 189
 2010, Hispanic, Male, Diabetes Mellitus, 187
 2010, White, Male, Intentional Self-Harm (Suicide), 182
 2010, Black, Male, Essential Hypertension and Renal Diseases, 178
 2010, Black, Male, Cerebrovascular Disease (stroke), 171
 2010, Hispanic, Female, Cerebrovascular Disease (stroke), 168
 2010, White, Female, Accidents Except Drug Poisoning, 163
 2010, Hispanic, Female, Chronic Lower Respiratory Diseases, 156
 2010, White, Male, Essential Hypertension and Renal Diseases, 153
 2010, Black, Female, Human Immunodeficiency Virus (HIV) Disease, 152
 2010, Hispanic, Male, Chronic Liver Disease and Cirrhosis, 150
 2010, Hispanic, Male, Chronic Lower Respiratory Diseases, 149
 2010, Black, Male, Accidents Except Drug Poisoning, 147
 2010, White, Male, Non-Hodgkins's Lymphoma, 141
 2010, Hispanic, Male, Cerebrovascular Disease (stroke), 140
 2010, White, Female, Leukemia, 139
 2010, Hispanic, Male, Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco, 138
 2010, Hispanic, Male, Accidents Except Drug Poisoning, 132
 2010, White, Female, Non-Hodgkins's Lymphoma, 127
 2010, Hispanic, Male, Assault (Homicide), 120
 2010, White, Male, Nephritis and Nephrotic Syndromes, 119
 2010, Black, Male, Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco, 111
 2010, White, Male, Chronic Liver Disease and Cirrhosis, 110
 2010, Asian, Male, Influenza (Flu) and Pneumonia, 108
 2010, White, Female, Septicemia, 107
 2010, Hispanic, Female, Essential Hypertension and Renal Diseases, 95
 2010, White, Male, Alzheimer's Disease, 94
 2010, White, Male, Insitu or Benign / Uncertain Neoplasms, 91
 2010, White, Female, Nephritis and Nephrotic Syndromes, 90
 2010, Asian, Female, Cerebrovascular Disease (stroke), 89
 2010, Unknown, Male, Diseases of Heart, 88
 2010, Hispanic, Male, Viral Hepatitis, 88
 2010, Black, Female, Alzheimer's Disease, 88
 2010, Unknown, Female, Diseases of Heart, 86
 2010, Black, Female, Nephritis and Nephrotic Syndromes, 86
 2010, Asian, Male, Cerebrovascular Disease (stroke), 83
 2010, White, Male, Viral Hepatitis, 82
 2010, White, Male, Parkinson's Disease, 79
 2010, Hispanic, Female, Human Immunodeficiency Virus (HIV) Disease, 79
 2010, White, Male, Human Immunodeficiency Virus (HIV) Disease, 76
 2010, White, Female, Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco, 75
 2010, Hispanic, Male, Intentional Self-Harm (Suicide), 75
 2010, Hispanic, Male, Essential Hypertension and Renal Diseases, 75
 2010, White, Male, Septicemia, 74
 2010, White, Female, Atherosclerosis, 74
 2010, Asian, Female, Influenza (Flu) and Pneumonia, 74
 2010, Hispanic, Female, Alzheimer's Disease, 71
 2010, Black, Male, Nephritis and Nephrotic Syndromes, 70
 2010, Black, Male, Certain Conditions Originating in the Perinatal Period, 70
 2010, White, Female, Insitu or Benign / Uncertain Neoplasms, 69
 2010, Black, Male, Chronic Liver Disease and Cirrhosis, 68
 2010, White, Female, Chronic Liver Disease and Cirrhosis, 67
 2010, Black, Female, Certain Conditions Originating in the Perinatal Period, 66
 2010, Black, Male, Intentional Self-Harm (Suicide), 64
 2010, Asian, Male, Diabetes Mellitus, 64
 2010, Black, Female, Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco, 63
 2010, White, Female, Intentional Self-Harm (Suicide), 61
 2010, Asian, Male, Chronic Lower Respiratory Diseases, 61
 2010, Hispanic, Female, Accidents Except Drug Poisoning, 58
 2010, White, Male, Mental and Behavioral Disorders due to use of Alcohol, 57
 2010, White, Female, Parkinson's Disease, 55
 2010, Unknown, Male, Malignant Neoplasms (cancer), 55
 2010, Black, Female, Septicemia, 55
 2010, Hispanic, Male, Certain Conditions Originating in the Perinatal Period, 54
 2010, Asian, Male, Accidents Except Drug Poisoning, 54
 2010, Asian, Female, Diabetes Mellitus, 51
 2010, Hispanic, Female, Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco, 50
 2010, Black, Male, Mental and Behavioral Disorders due to use of Alcohol, 50
 2010, White, Male, Aortic Aneurysm and Dissection, 49
 2010, Hispanic, Female, Chronic Liver Disease and Cirrhosis, 49
 2010, Black, Male, Viral Hepatitis, 49
 2010, Black, Male, Leukemia, 49
 2010, Hispanic, Female, Nephritis and Nephrotic Syndromes, 48
 2010, Black, Female, Non-Hodgkins's Lymphoma, 48
 2010, Black, Female, Leukemia, 47
 2010, White, Male, Atherosclerosis, 46
 2010, Hispanic, Male, Mental and Behavioral Disorders due to use of Alcohol, 46
 2010, Asian, Male, Intentional Self-Harm (Suicide), 46
 2010, Hispanic, Male, Leukemia, 45
 2010, Black, Female, Assault (Homicide), 45
 2010, Hispanic, Male, Nephritis and Nephrotic Syndromes, 43
 2010, Black, Female, Chronic Liver Disease and Cirrhosis, 43
 2010, Hispanic, Female, Leukemia, 42
 2010, Hispanic, Female, Congenital Malformations and Chromosomal Abnormalities, 42
 2010, Hispanic, Female, Viral Hepatitis, 40
 2010, Hispanic, Female, Non-Hodgkins's Lymphoma, 40
 2010, Black, Male, Non-Hodgkins's Lymphoma, 39
 2010, Hispanic, Male, Congenital Malformations and Chromosomal Abnormalities, 38

2010,Hispanic,Female,Certain Conditions Originating in the Perinatal Period,38
 2010,Black,Male,Septicemia,36
 2010,White,Male,Certain Conditions Originating in the Perinatal Period,35
 2010,Unknown,Female,Malignant Neoplasms (cancer),35
 2010,Hispanic,Male,Non-Hodgkins's Lymphoma,35
 2010,Black,Male,Congenital Malformations and Chromosomal Abnormalities,34
 2010,White,Male,Congenital Malformations and Chromosomal Abnormalities,33
 2010,Hispanic,Female,Septicemia,33
 2010,Black,Female,Viral Hepatitis,33
 2010,Black,Female,Congenital Malformations and Chromosomal Abnormalities,33
 2010,Asian,Female,Essential Hypertension and Renal Diseases,33
 2010,White,Female,Viral Hepatitis,32
 2010,White,Female,Congenital Malformations and Chromosomal Abnormalities,32
 2010,White,Female,Certain Conditions Originating in the Perinatal Period,31
 2010,Hispanic,Male,Septicemia,31
 2010,Hispanic,Male,Alzheimer's Disease,31
 2010,Asian,Male,Essential Hypertension and Renal Diseases,30
 2010,Hispanic,Male,Aortic Aneurysm and Dissection,28
 2010,Asian,Female,Chronic Lower Respiratory Diseases,28
 2010,Hispanic,Female,Intentional Self-Harm (Suicide),27
 2010,Black,Female,Insitu or Benign / Uncertain Neoplasms,26
 2010,White,Male,Assault (Homicide),25
 2010,Black,Male,Aortic Aneurysm and Dissection,25
 2010,Asian,Female,Accidents Except Drug Poisoning,25
 2010,White,Female,Human Immunodeficiency Virus (HIV) Disease,24
 2010,Hispanic,Female,Insitu or Benign / Uncertain Neoplasms,23
 2010,Black,Female,Atherosclerosis,23
 2010,White,Female,Peptic Ulcer,22
 2010,White,Female,Aortic Aneurysm and Dissection,22
 2010,Hispanic,Female,Assault (Homicide),22
 2010,Black,Male,Atherosclerosis,20
 2010,Black,Male,Anemias,20
 2010,Black,Male,Alzheimer's Disease,20
 2010,White,Male,Peptic Ulcer,19
 2010,White,Female,Cholelithiasis and Disorders of Gallbladder,19
 2010,Black,Male,Insitu or Benign / Uncertain Neoplasms,19
 2010,Black,Female,Intentional Self-Harm (Suicide),19
 2010,Black,Female,Anemias,19
 2010,Asian,Male,Non-Hodgkins's Lymphoma,19
 2010,Hispanic,Male,Parkinson's Disease,18
 2010,Hispanic,Male,Insitu or Benign / Uncertain Neoplasms,18
 2010,Black,Female,Pregnancy and Childbirth,18
 2010,Asian,Male,Chronic Liver Disease and Cirrhosis,18
 2010,Asian,Female,Certain Conditions Originating in the Perinatal Period,18

2010,Asian,Male,Leukemia,17			
2010,Asian,Male,Certain Conditions Originating in the Perinatal Period,17			
2010,Asian,Female,Intentional Self-Harm (Suicide),17			
2010,White,Female,Mental and Behavioral Disorders due to use of Alcohol,16			
2010,Hispanic,Female,Parkinson's Disease,16			
2010,Black,Female,Aortic Aneurysm and Dissection,16			
2010,Unknown,Male,Chronic Lower Respiratory Diseases,15			
2010,Hispanic,Female,Atherosclerosis,15			
2010,Black,Female,Peptic Ulcer,15			
2010,Asian,Female,Nephritis and Nephrotic Syndromes,15			
2010,Black,Female,Maternal Causes,14			
2010,Asian,Male,Viral Hepatitis,14			
2010,Asian,Female,Alzheimer's Disease,14			
2010,White,Female,Assault (Homicide),13			
2010,Hispanic,Male,Atherosclerosis,13			
2010,Black,Male,Peptic Ulcer,13			
2010,Asian,Female,Non-Hodgkins's Lymphoma,13			
2010,Hispanic,Female,Pregnancy and Childbirth,11			
2010,Black,Female,Mental and Behavioral Disorders due to use of Alcohol,11			
2010,Black,Female,Complications of Medical and Surgical Care,11			
2010,Asian,Male,Septicemia,11			
2010,Asian,Male,Nephritis and Nephrotic Syndromes,11			
2010,Asian,Female,Leukemia,11			
2010,White,Male,Cholelithiasis and Disorders of Gallbladder,10			
2010,Unknown,Male,Accidents Except Drug Poisoning,10			
2010,Hispanic,Male,Anemias,10			
2010,Hispanic,Female,Cholelithiasis and Disorders of Gallbladder,10			
2010,Asian,Female,Chronic Liver Disease and Cirrhosis,10			
2010,Hispanic,Male,Peptic Ulcer,9			
2010,Hispanic,Female,Peptic Ulcer,9			
2010,Hispanic,Female,Maternal Causes,9			
2010,Black,Female,Parkinson's Disease,9			
2010,Asian,Male,Mental and Behavioral Disorders due to use of Alcohol,9			
2010,Asian,Male,Congenital Malformations and Chromosomal Abnormalities,9			
2010,Asian,Male,Assault (Homicide),9			
2010,Asian,Male,Alzheimer's Disease,9			
2010,Asian,Female,Insitu or Benign / Uncertain Neoplasms,9			
2010,White,Male,Pneumonitis due to Solids and Liquids,8			
2010,White,Male,Anemias,8			
2010,Unknown,Male,Human Immunodeficiency Virus (HIV) Disease,8			
2010,Unknown,Male,Diabetes Mellitus,8			
2010,Hispanic,Male,Tuberculosis (TB),8			
2010,White,Male,Complications of Medical and Surgical Care,7			
2010,Black,Male,Parkinson's Disease,7			

2010,Black,Female,Cholelithiasis and Disorders of Gallbladder,7
 2010,Asian,Male,Parkinson's Disease,7
 2010,Asian,Male,Insitu or Benign / Uncertain Neoplasms,7
 2010,Asian,Female,Septicemia,7
 2010,Asian,Female,Congenital Malformations and Chromosomal Abnormalities,7
 2010,Asian,Female,Assault (Homicide),7
 2010,White,Female,Anemias,6
 2010,Unknown,Male,Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco,6
 2010,Unknown,Male,Influenza (Flu) and Pneumonia,6
 2010,Unknown,Male,Certain Conditions Originating in the Perinatal Period,6
 2010,Unknown,Male,Assault (Homicide),6
 2010,Unknown,Female,Diabetes Mellitus,6
 2010,Unknown,Female,Chronic Lower Respiratory Diseases,6
 2010,Unknown,Female,Cerebrovascular Disease (stroke),6
 2010,Hispanic,Female,Mental and Behavioral Disorders due to use of Alcohol,6
 2010,Black,Female,Meningitis,6
 2010,Asian,Male,Cholelithiasis and Disorders of Gallbladder,6
 2010,Asian,Male,Atherosclerosis,6
 2010,White,Female,Pneumonitis due to Solids and Liquids,5
 2010,White,Female,Complications of Medical and Surgical Care,5
 2010,Unknown,Male,Viral Hepatitis,5
 2010,Unknown,Male,Mental and Behavioral Disorders due to use of Alcohol,5
 2010,Unknown,Male,Intentional Self-Harm (Suicide),5
 2010,Unknown,Male,Chronic Liver Disease and Cirrhosis,5
 2010,Unknown,Female,Essential Hypertension and Renal Diseases,5
 2010,Unknown,Female,Certain Conditions Originating in the Perinatal Period,5
 2010,Other,Male,Accidents Except Drug Poisoning,5
 2010,Hispanic,Male,Cholelithiasis and Disorders of Gallbladder,5
 2010,Hispanic,Female,Aortic Aneurysm and Dissection,5
 2010,Black,Male,Pneumonitis due to Solids and Liquids,5
 2010,Asian,Male,Tuberculosis (TB),5
 2010,Asian,Male,Aortic Aneurysm and Dissection,5
 2010,Asian,Male,Anemias,5
 2010,Asian,Female,Viral Hepatitis,5
 2010,Unknown,Male,Cerebrovascular Disease (stroke),4
 2010,Unknown,Female,Influenza (Flu) and Pneumonia,4
 2010,Hispanic,Female,Meningitis,4
 2010,Hispanic,Female,Complications of Medical and Surgical Care,4
 2010,Black,Male,Meningitis,4
 2010,Black,Male,Cholelithiasis and Disorders of Gallbladder,4
 2010,Black,Female,Tuberculosis (TB),4
 2010,Black,Female,Pneumonitis due to Solids and Liquids,4
 2010,Asian,Male,Peptic Ulcer,4
 2010,Asian,Male,Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco,4

```

2010,Asian,Female,Pregnancy and Childbirth,4
2010,Asian,Female,Peptic Ulcer,4
2010,Asian,Female,Parkinson's Disease,4
2010,Asian,Female,Maternal Causes,4
2010,Asian,Female,Atherosclerosis,4
2010,Asian,Female,Aortic Aneurysm and Dissection,4
2010,White,Male,Tuberculosis (TB),3
2010,White,Male,Meningitis,3
2010,White,Female,Meningitis,3
2010,Unknown,Male,Essential Hypertension and Renal Diseases,3
2010,Unknown,Female,Viral Hepatitis,3
2010,Unknown,Female,Insitu or Benign / Uncertain Neoplasms,3
2010,Other,Female,Accidents Except Drug Poisoning,3
2010,Hispanic,Male,Meningitis,3
2010,Hispanic,Male,Legal Intervention,3
2010,Black,Male,Complications of Medical and Surgical Care,3
2010,Asian,Male,Human Immunodeficiency Virus (HIV) Disease,3
2010,Asian,Female,Human Immunodeficiency Virus (HIV) Disease,3
2010,White,Female,Pregnancy and Childbirth,2
2010,White,Female,Maternal Causes,2
2010,White,Female,All Censored Causes,2
2010,Unknown,Male,Peptic Ulcer,2
2010,Unknown,Male,Nephritis and Nephrotic Syndromes,2
2010,Unknown,Female,Septicemia,2
2010,Unknown,Female,Nephritis and Nephrotic Syndromes,2
2010,Unknown,Female,Atherosclerosis,2
2010,Unknown,Female,Accidents Except Drug Poisoning,2
2010,Hispanic,Male,Pneumonitis due to Solids and Liquids,2
2010,Hispanic,Male,All Censored Causes,2
2010,Hispanic,Female,Tuberculosis (TB),2
2010,Hispanic,Female,Pneumonitis due to Solids and Liquids,2
2010,Black,Male,Tuberculosis (TB),2
2010,Black,Male,All Censored Causes,2
2010,Black,Female,All Censored Causes,2
2010,Asian,Female,Cholelithiasis and Disorders of Gallbladder,2
2010,Asian,Female,All Censored Causes,2
2010,White,Male,Legal Intervention,1
2010,White,Male,All Censored Causes,1
2010,White,Female,Tuberculosis (TB),1
2010,Unknown,Male,Septicemia,1
2010,Unknown,Male,Leukemia,1
2010,Unknown,Male,Complications of Medical and Surgical Care,1
2010,Unknown,Male,Atherosclerosis,1
2010,Unknown,Male,Aortic Aneurysm and Dissection,1
2010,Unknown,Female,Pregnancy and Childbirth,1
2010,Unknown,Female,Parkinson's Disease,1
2010,Unknown,Female,Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco,1
2010,Unknown,Female,Maternal Causes,1
2010,Unknown,Female,Leukemia,1
2010,Unknown,Female,Intentional Self-Harm (Suicide),1
2010,Unknown,Female,Alzheimer's Disease,1
2010,Hispanic,Male,Complications of Medical and Surgical Care,1
2010,Hispanic,Female,Anemias,1
2010,Hispanic,Female,All Censored Causes,1
2010,Black,Male,Legal Intervention,1
2010,Asian,Male,Meningitis,1
2010,Asian,Male,Legal Intervention,1
2010,Asian,Male,Complications of Medical and Surgical Care,1
2010,Asian,Male,All Censored Causes,1
2010,Asian,Female,Tuberculosis (TB),1
2010,Asian,Female,Mental and Behavioral Disorders due to use of or Accidental Poisoning by Psychoactive Substances Excluding Alcohol and Tobacco,1
2010,Asian,Female,Anemias,1
~/REPOS/usd-ads/ads500b/module02 >>>

```

The top three causes of death for Black Males:

```

~/REPOS/usd-ads/ads500b/module02 >>> grep Black Exercise1_data.csv | grep Male | sort -t, -r -n -k5,5
2010,Black,Male,Diseases of Heart,2015
2010,Black,Male,Malignant Neoplasms (cancer),1540
2010,Black,Male,Assault (Homicide),299
2010,Black,Male,Human Immunodeficiency Virus (HIV) Disease,207

```

The bottom five causes of death for Hispanic Females

```
~/REPOS/usd-ads/ads500b/module02 >>> grep Hispanic Exercise1_data.csv | grep Female | sort -t, -n -k5,5
2010,Hispanic,Female,All Censored Causes,1
2010,Hispanic,Female,Anemias,1
2010,Hispanic,Female,Pneumonitis due to Solids and Liquids,2
2010,Hispanic,Female,Tuberculosis (TB),2
2010,Hispanic,Female,Complications of Medical and Surgical Care,4
```

Exercise Two

Obtained from UNICEF, the Exercise 2 Dataset (located in your assignment prompt in Blackboard) 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?

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

Answer 2.1

```
~/REPOS/usd-ads/ads500b/module02 >>> cat Exercise2_data.csv | sort -t, -n -k4,4
Burundi,10864245,1380411,13
```

Answer 2.2

```
~/REPOS/usd-ads/ads500b/module02 >>> awk -F, '{if ($3 > 10000000 && $4 < 50) {print $1,$2,$3,$4 } }' Exercise2_data.csv
Bangladesh 164669750 59047282 36
Democratic Republic of the Congo 81339984 35691983 44
Egypt 97553148 41660074 43
Ethiopia 104957438 21316855 20
India 1339180125 449964502 34
Kenya 49699863 13201277 27
Myanmar 53370609 16183036 30
Pakistan 197015953 71796556 36
Philippines 104918094 48977864 47
Sudan 40533328 13931304 34
Thailand 69037516 33966458 49
United Republic of Tanzania 57310020 18942682 33
Viet Nam 95540797 33642782 35
Yemen 28250420 10174671 36
~/REPOS/usd-ads/ads500b/module02 >>>
```


Exercise Three

For the following exercise, use the Exercise 3 Dataset (located in your assignment prompt in Blackboard), which contains 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

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

3.2: List the top five countries with the highest percentage of public and private median availability of selected medicines in 2007–2013.

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.

Answer 3.1

```
~/REPOS/usd-ads/ads500b/module02 >>> cat Exercise3_data.csv | tr -d "\"" | grep -v Country | grep -v Median | sort -t, -n -k2,2
India,2.8,
```

Answer 3.2

```
~/REPOS/usd-ads/ads500b/module02 >>> cat Exercise3_data.csv | tr -d "\"" | grep -v Country | grep -v Median | sort -t, -r -n -k2,2 -k3,3
Russian Federation,100.0,100.0
Syrian Arab Republic,98.2,93.0
Iran (Islamic Republic of),96.7,96.7
Afghanistan,94.0,81.1
Sudan,91.7,77.1
```


Answer 3.3

```
~/Repos/usd-ads/ads500b/module02 >>> cat Exercise3_data.csv | tr -d "\"" | grep -v Country | grep -v Median | awk -F, '{if($2>$3) print $0","($2-$3)}' | sort -t, -n -r -k4,4
Brazil,76.7,0.0,76.7
Kyrgyzstan,70.0,,70
Bolivia (Plurinational State of),86.7,31.9,54.8
```

Reasons for selecting this answer:

- 1 The gap between the availability in the private and public sectors is 54.0 percent and up to 76.7 percent.
- 2 Typically, the private sector offers better service.

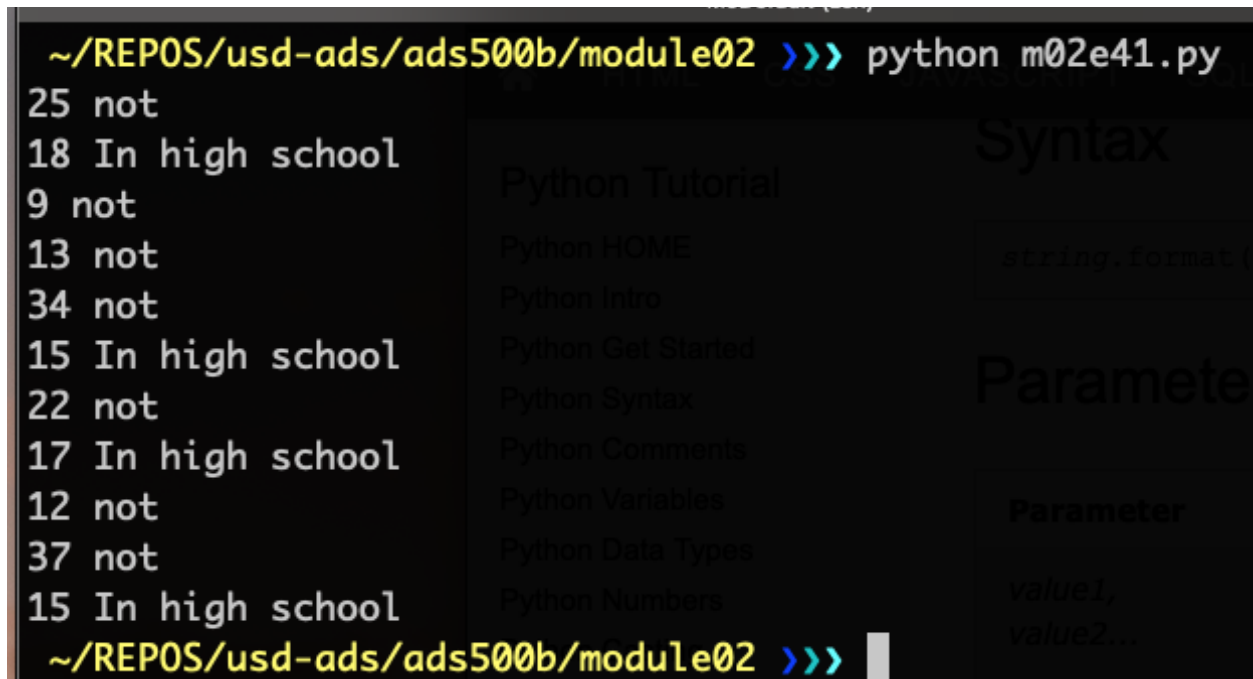
Exercise Four

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.

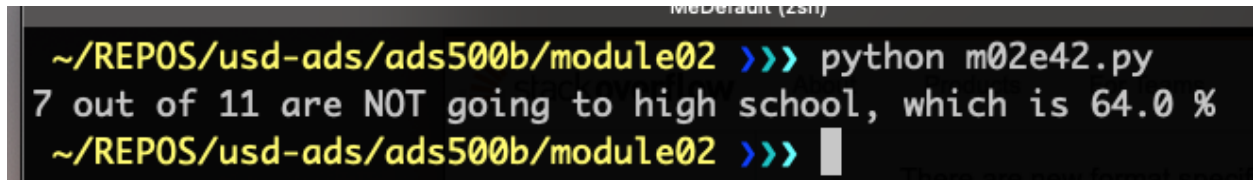
4.2: From the list, calculate the percentage of people not going to high school.

Answers 4.1



```
~/REPOS/usd-ads/ads500b/module02 >>> python m02e41.py
25 not
18 In high school
9 not
13 not
34 not
15 In high school
22 not
17 In high school
12 not
37 not
15 In high school
~/REPOS/usd-ads/ads500b/module02 >>>
```

For the code, please see Figure 1 in the appendix.

Answer 4.2A terminal window with a dark background. The prompt is a yellow path: ~/REPOS/usd-ads/ads500b/module02. The user enters a command with three blue chevrons followed by python m02e42.py. The output is white text: 7 out of 11 are NOT going to high school, which is 64.0 %. The prompt is repeated on the next line with a white cursor bar at the end.

```
~/REPOS/usd-ads/ads500b/module02 >>> python m02e42.py
7 out of 11 are NOT going to high school, which is 64.0 %
~/REPOS/usd-ads/ads500b/module02 >>> █
```

For the code, please see Figure 2 in the appendix.

Appendix

Figure 1

Code for Answer 4.1

```
# ADS500B - Module 04 - Exercise 4.1
# Mayel Espino
# University of San Diego - Applied Data Science
age = [25, 18, 9, 13, 34, 15, 22, 17, 12, 37, 15]
for a in age:
    if a >= 14 and a <= 18:
        print a, "In high school"
    else:
        print a, "not"
~
```

Figure 2

Code for Answer 4.2

```
# ADS500B - Module 04 - Exercise 4.2
# Mayel Espino
# University of San Diego - Applied Data Science
age = [25, 18, 9, 13, 34, 15, 22, 17, 12, 37, 15]
hscount = 0
for a in age:
    if a < 14 or a > 18:
        hscount += 1
print "{0} out of {1} are NOT going to high school, which is {2} %".format(hscount, len(age), round((float(hscount)/len(age))*100))
~
```

References