

EE114 Coding HW 3

1) *Plot the histogram of continuous features*

This part is to show how the population is distributed among certain categories

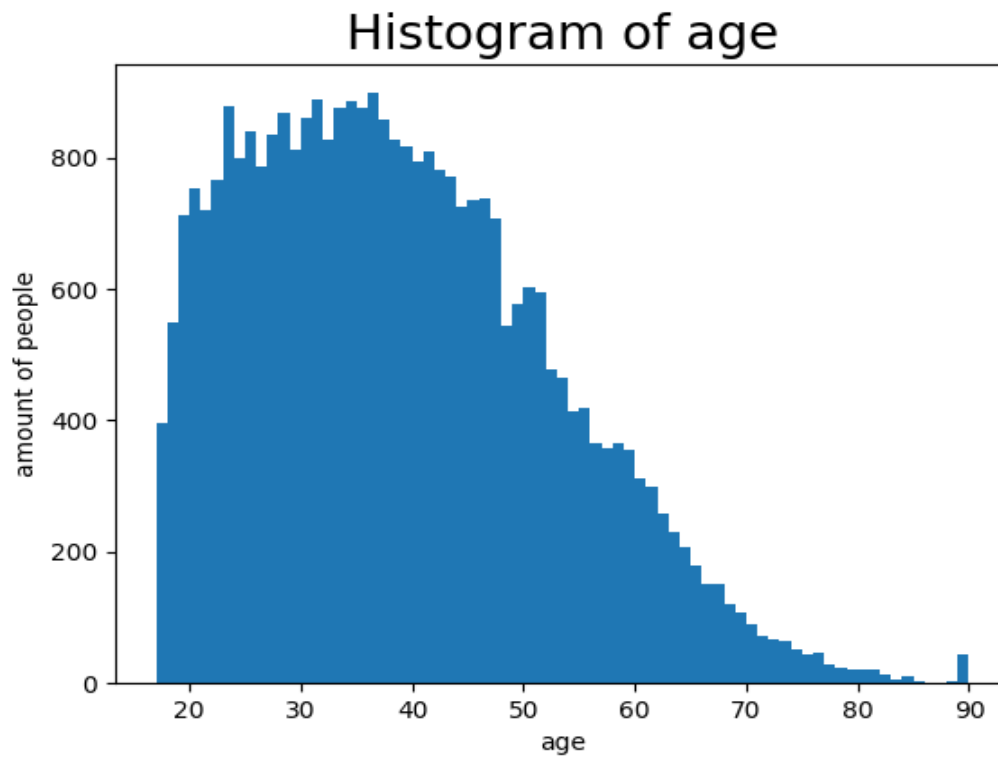


Figure 1 Histogram of age

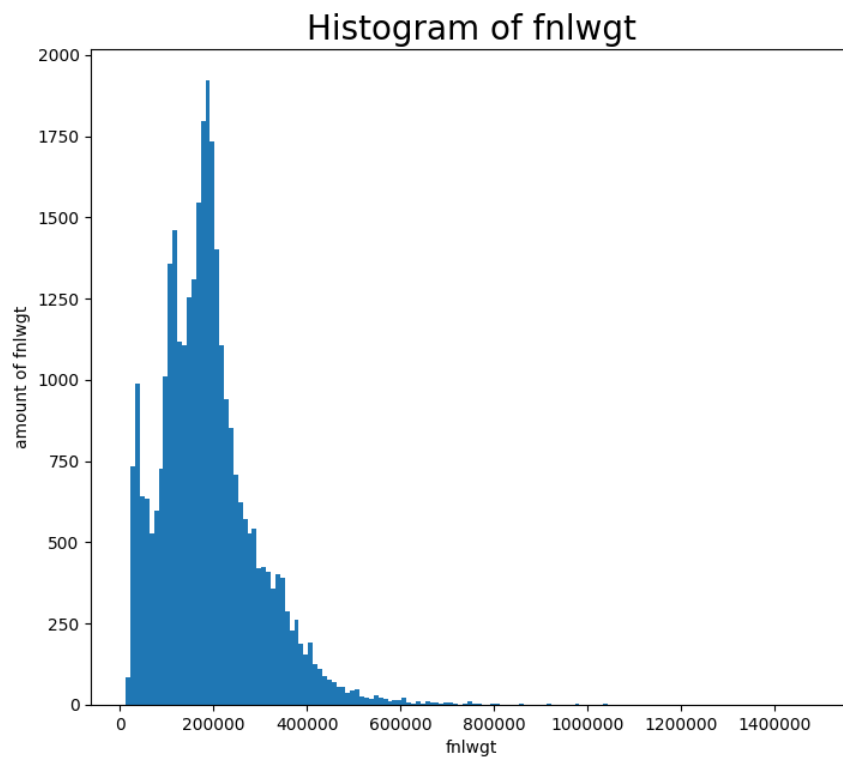


Figure 2 Histogram of fnlwgt

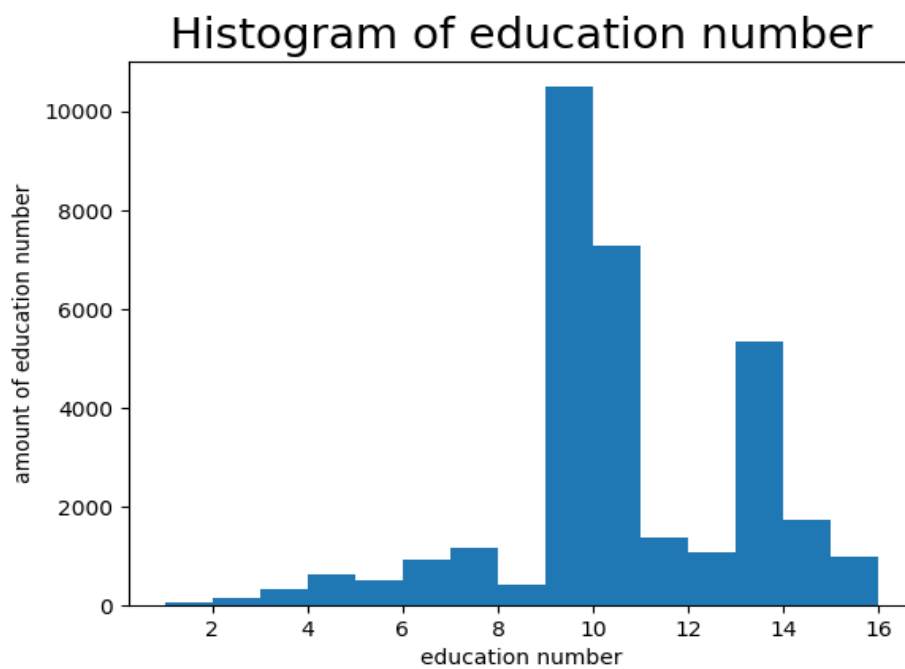


Figure 3 Hist of educatin number

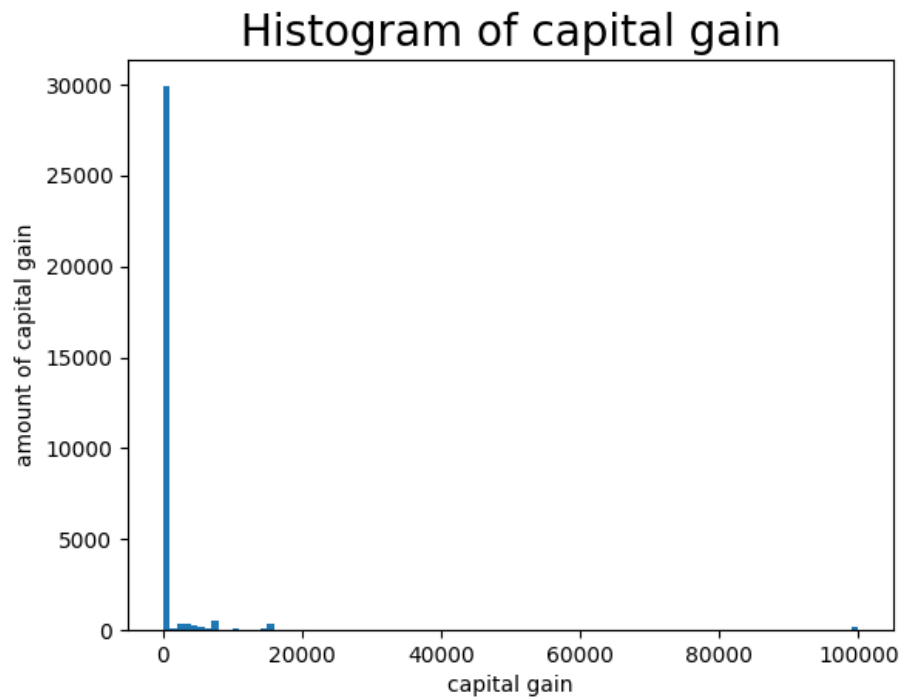


Figure 4 Histogram of captical gain

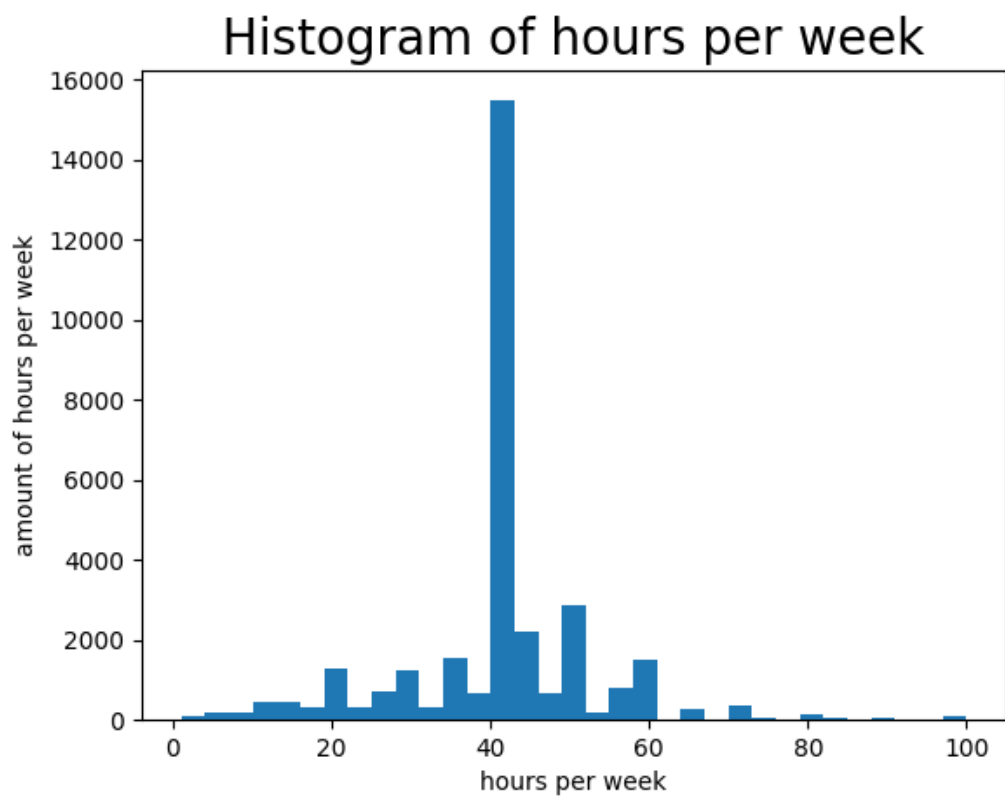


Figure 5 Histogram of hours per week

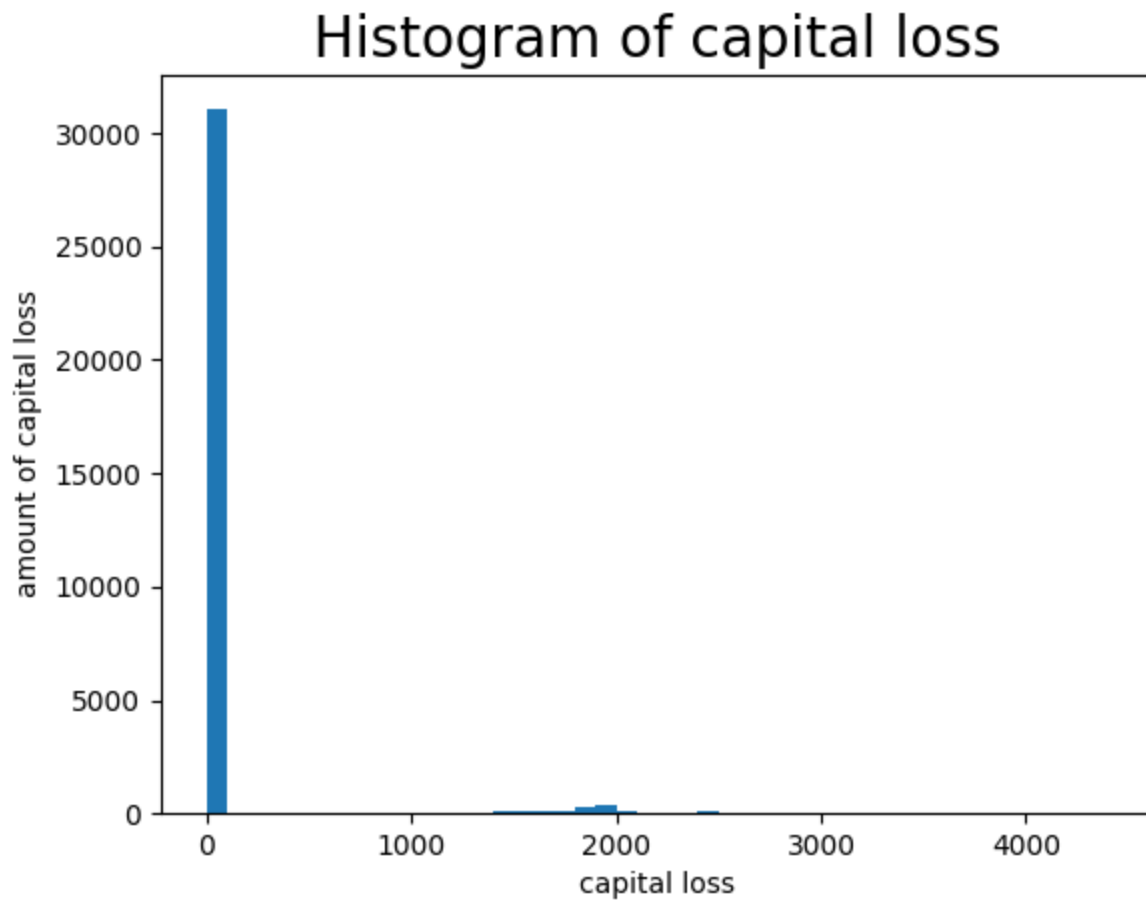


Figure 6 Histogram of capital loss

2) Calculate the correlation coefficient between income and each of the continuous features.

```
correlation between age and income is 0.23403710264878327
correlation between fnlwgt and income is -0.009462557247526213
correlation between education number and income is 0.335153952690807
correlation between capital gain and income is 0.22332881819526118
correlation between capital loss and income is 0.1505263117702336
correlation between hours per week and income is 0.229689065670786
```

Figure 7 Correlation Coefficient of continuous and income

3) Calculate the correlation coefficient between income and each value of each discrete feature. Discrete features are workclass, education, marital-status, relationship, and sex. For instance, for workclass, you need to go over the values "Private, Self-emp-not-inc, Self-emp-inc, ..." and look at the coefficient between income and 0/1 vector obtained by that workclass value as discussed above.

```

CORRELATION OF workclass

correlation between income and Private is -0.07852827453627606
correlation between income and Self-emp-not-inc is 0.030019053961932152
correlation between income and Self-emp-inc is 0.1394665573405462
correlation between income and Federal-gov is 0.059371745431774675
correlation between income and Local-gov is 0.03309057154412962
correlation between income and State-gov is 0.014843290616531895
correlation between income and Without-pay is -0.011680731190572053
correlation between income and Never-worked is -0.008258636174571447

CORRELATION OF education

correlation between income and Bachelors is 0.18048542920208524
correlation between income and Some-college is -0.0635352554294096
correlation between income and 11th is -0.08586392327835302
correlation between income and HS-grad is -0.13118854401273908
correlation between income and Prof-school is 0.15490683578942085
correlation between income and Assoc-acdm is 0.003250289246823411
correlation between income and Assoc-voc is 0.010047780081320302
correlation between income and 9th is -0.05576743851103557
correlation between income and 7th-8th is -0.0595240949837756
correlation between income and 12th is -0.04469017135415736
correlation between income and Masters is 0.17456971065150761
correlation between income and 1st-4th is -0.03454398828996042
correlation between income and 10th is -0.0700378408728454
correlation between income and Doctorate is 0.13257200580959538
correlation between income and 5th-6th is -0.04582616682401098
correlation between income and Preschool is -0.022306842300548947

CORRELATION OF marital status

correlation between income and Married-civ-spouse is 0.4446961550459976
correlation between income and Divorced is -0.1269949961258107
correlation between income and Never-married is -0.3184403250735543
correlation between income and Separated is -0.07438622703110984
correlation between income and Widowed is -0.06438127591306757
correlation between income and Married-spouse-absent is -0.04253155700132926
correlation between income and Married-AF-spouse is 0.01206139838059884

CORRELATION OF occupation

correlation between income and Tech-support is 0.025696277739681656
correlation between income and Craft-repair is -0.01257565376099275
correlation between income and Other-service is -0.15634773905615337
correlation between income and Sales is 0.023688168354987382
correlation between income and Exec-managerial is 0.21486131464631295
correlation between income and Prof-specialty is 0.18586584592031466
correlation between income and Handlers-cleaners is -0.08726514883223992
correlation between income and Machine-op-inspct is -0.06940051324443093
correlation between income and Adm-clerical is -0.08998555784156023
correlation between income and Farming-fishing is -0.051924826461923895
correlation between income and Transport-moving is -0.02147616988216717
correlation between income and Priv-house-serv is -0.03712153884404855
correlation between income and Protective-serv is 0.028118465773859416
correlation between income and Armed-Forces is -0.00504376639928706

```

Figure 8 Correlation coefficient of discrete values

CORRELATION OF relationship

correlation between income and Wife is 0.1232635397280336
correlation between income and Own-child is -0.22853196347365237
correlation between income and Husband is 0.40103526357728686
correlation between income and Not-in-family is -0.1884973713038389
correlation between income and Other-relative is -0.08371649655545951
correlation between income and Unmarried is -0.1428572613718598

CORRELATION OF race

correlation between income and White is 0.08522449750183451
correlation between income and Asian-Pac-Islander is 0.010543232871471162
correlation between income and Amer-Indian-Eskimo is -0.028721059409003217
correlation between income and Other is -0.03182997497887226
correlation between income and Black is -0.0890886033221676

CORRELATION OF sex

correlation between income and Female is -0.21598015058412695
correlation between income and Male is 0.21598015058412695

CORRELATION OF native_country

correlation between income and United-States is 0.03447035976701251
correlation between income and Cambodia is 0.007211610318623418
correlation between income and England is 0.01139238215946561
correlation between income and Puerto-Rico is -0.01879061141576641
correlation between income and Canada is 0.011641838819113397
correlation between income and Germany is 0.012216467450162475
correlation between income and Outlying-US(Guam-USVI-etc) is -0.011680731190573553
correlation between income and India is 0.020664448393075973
correlation between income and Japan is 0.014943591843530628
correlation between income and Greece is 0.0024476568517178524
correlation between income and South is -0.004736744473148106
correlation between income and China is 0.002905696536757044
correlation between income and Cuba is 0.0028273553934038977
correlation between income and Iran is 0.015120977283996386
correlation between income and Honduras is -0.0076602055861275395
correlation between income and Philippines is 0.012306242941019344
correlation between income and Italy is 0.011269939123838892
correlation between income and Poland is -0.004100878692339336
correlation between income and Jamaica is -0.013706154044150971
correlation between income and Vietnam is -0.017648564101976447
correlation between income and Mexico is -0.06290081708893568
correlation between income and Portugal is -0.010467966549098415
correlation between income and Ireland is -0.002062861977656272
correlation between income and France is 0.012079144781716124
correlation between income and Dominican-Republic is -0.02303979779570602
correlation between income and Laos is -0.007133949115141156
correlation between income and Ecuador is -0.006720783095390512
correlation between income and Taiwan is 0.014019711292711621
correlation between income and Haiti is -0.012896203018829964
correlation between income and Columbia is -0.020617824869144992
correlation between income and Hungary is -0.00046929414952905926
correlation between income and Guatemala is -0.020128496174504316
correlation between income and Nicaragua is -0.013760787051011794
correlation between income and Scotland is 0.00041271095486759954
correlation between income and Thailand is -0.004078164581120435
correlation between income and Yugoslavia is 0.006958685561911
correlation between income and El-Salvador is -0.020838040188019196
correlation between income and Trinidad&Tobago is -0.00766002946862229
correlation between income and Peru is -0.012728074630025066

Figure 9 Continued Correlation coefficient of discrete values

4) For each discrete feature, create a bar plot where x-axis is the feature values (e.g. "Private, Self-emp-not-inc, Self-emp-inc, ... ") and y-axis is the correlation coefficients obtained in the previous question.

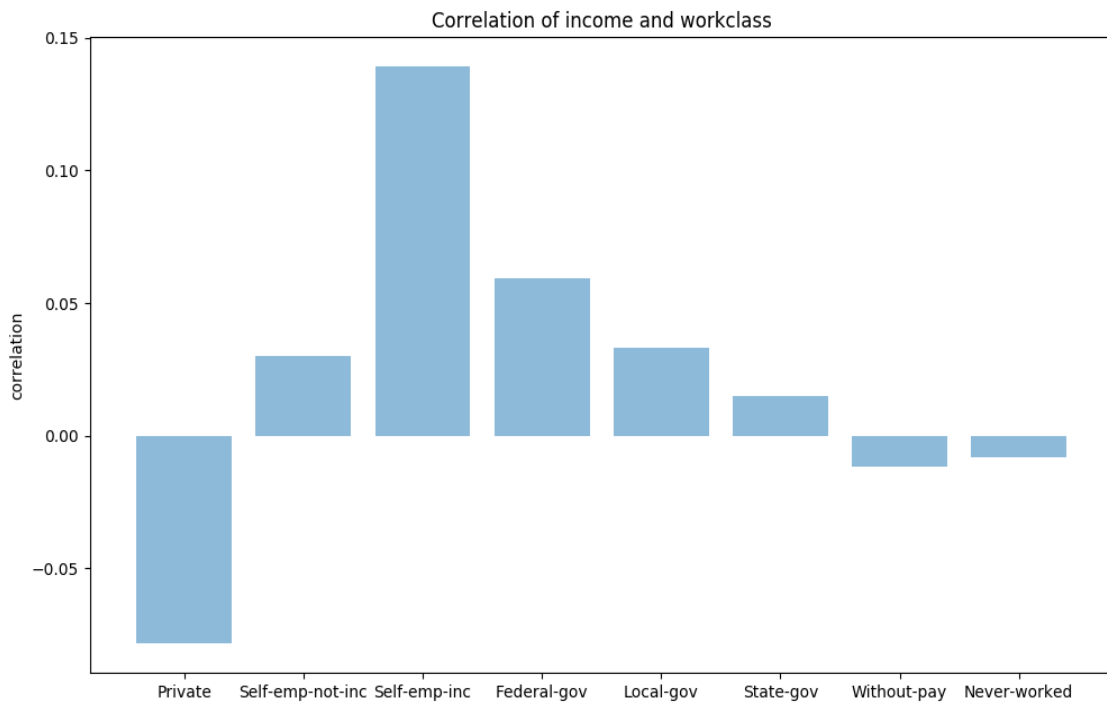


Figure 10 Bar graph of correlation coefficient of workclass divisions

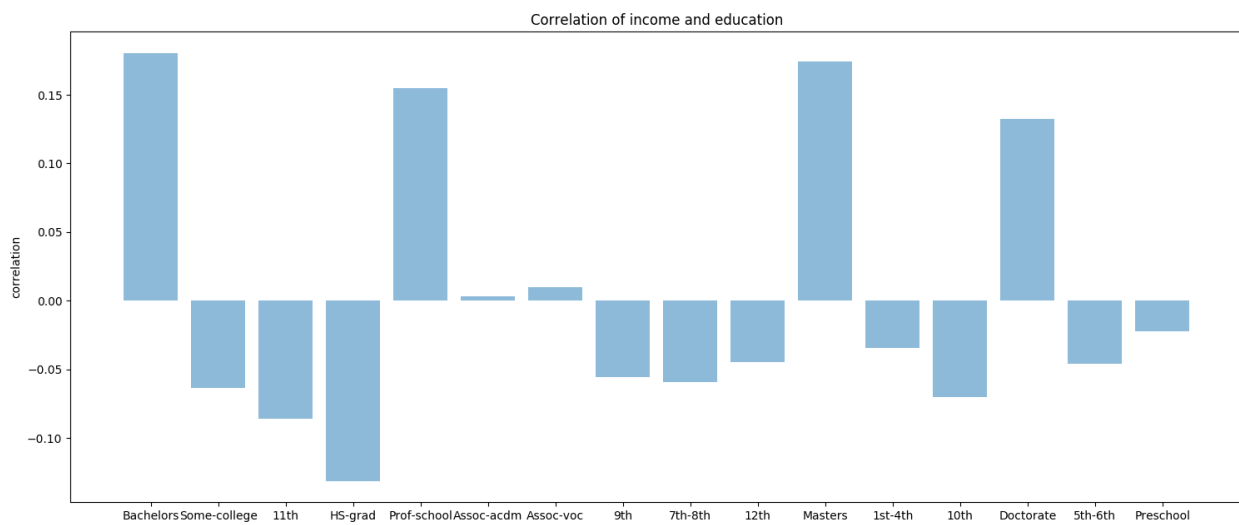


Figure 11 Bar graph of correlation coefficient of education divisions

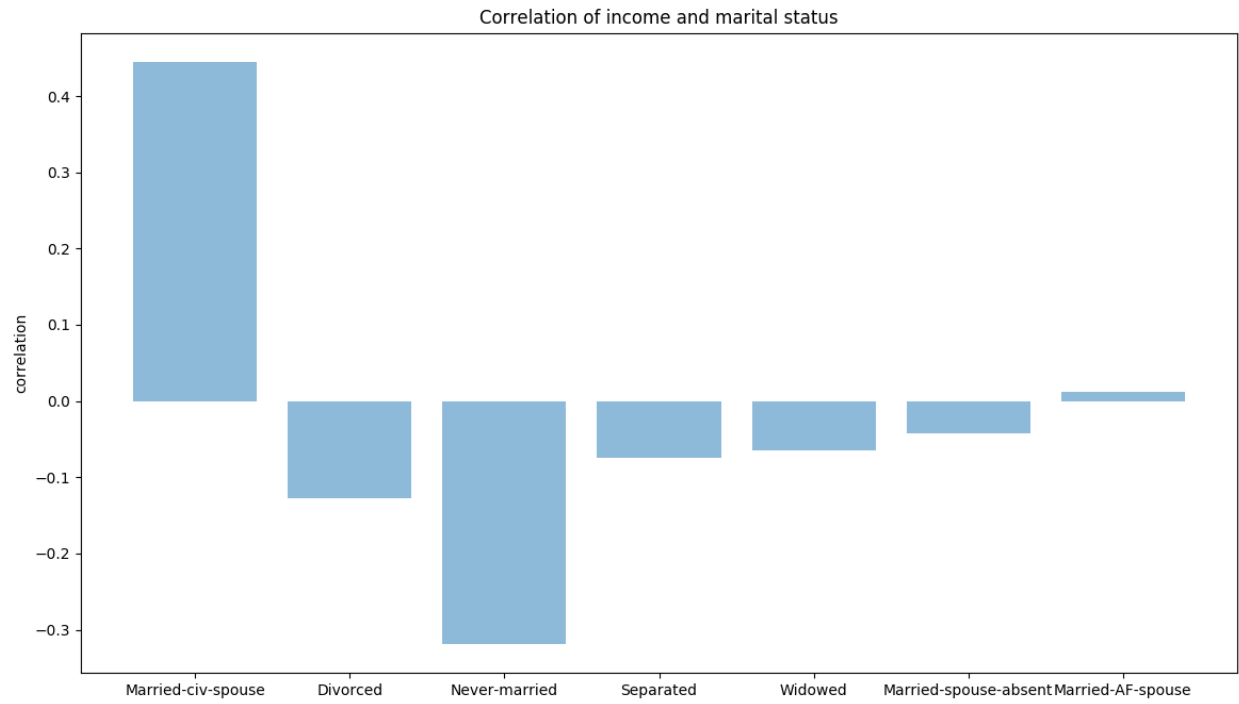


Figure 12 Bar graph of correlation coefficient of marital status divisions

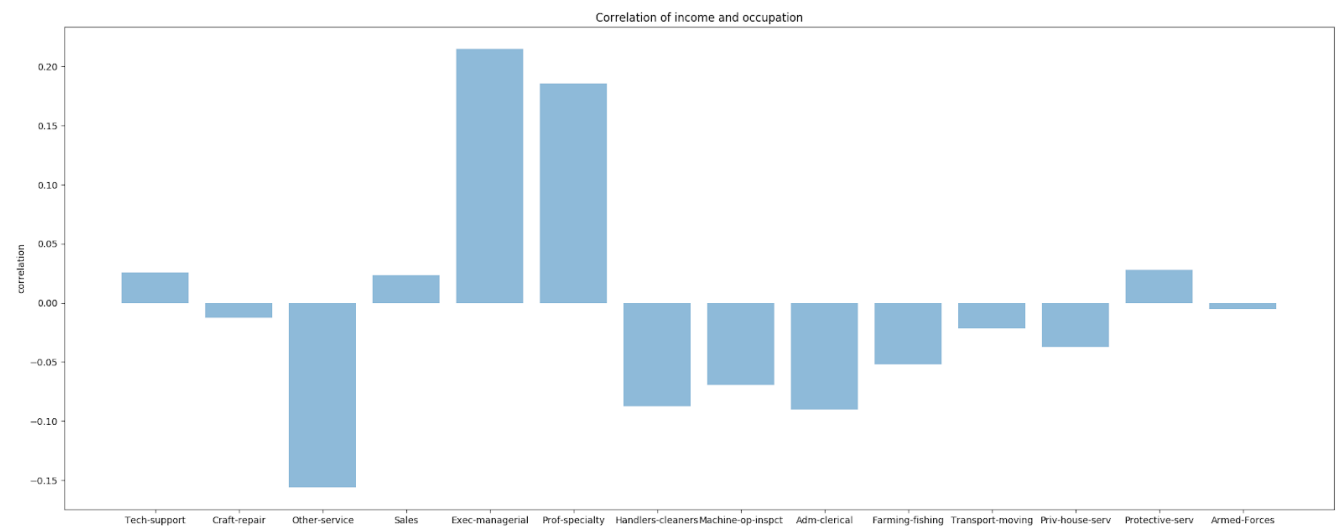


Figure 13 Bar graph of correlation coefficient of occupation divisions

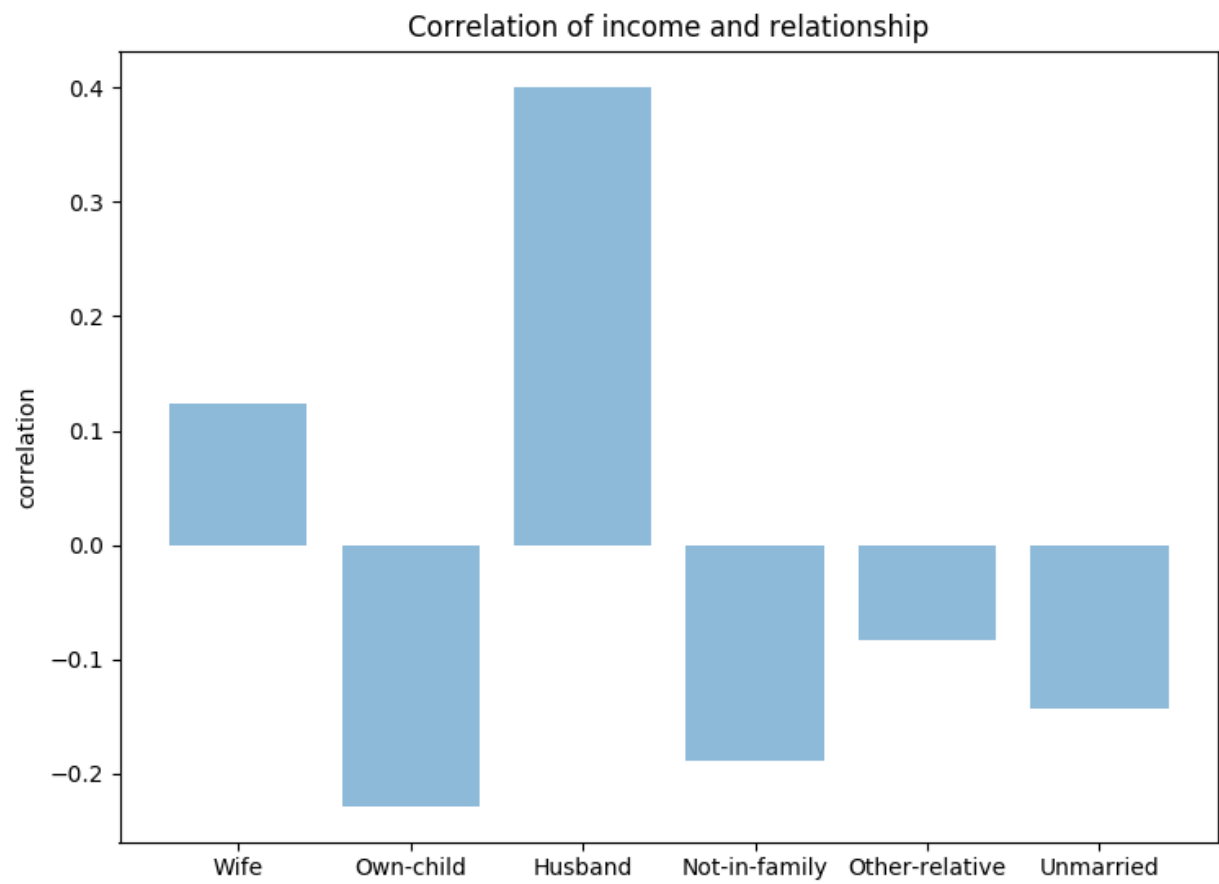


Figure 14 Bar graph of correlation coefficient of relationship divisions

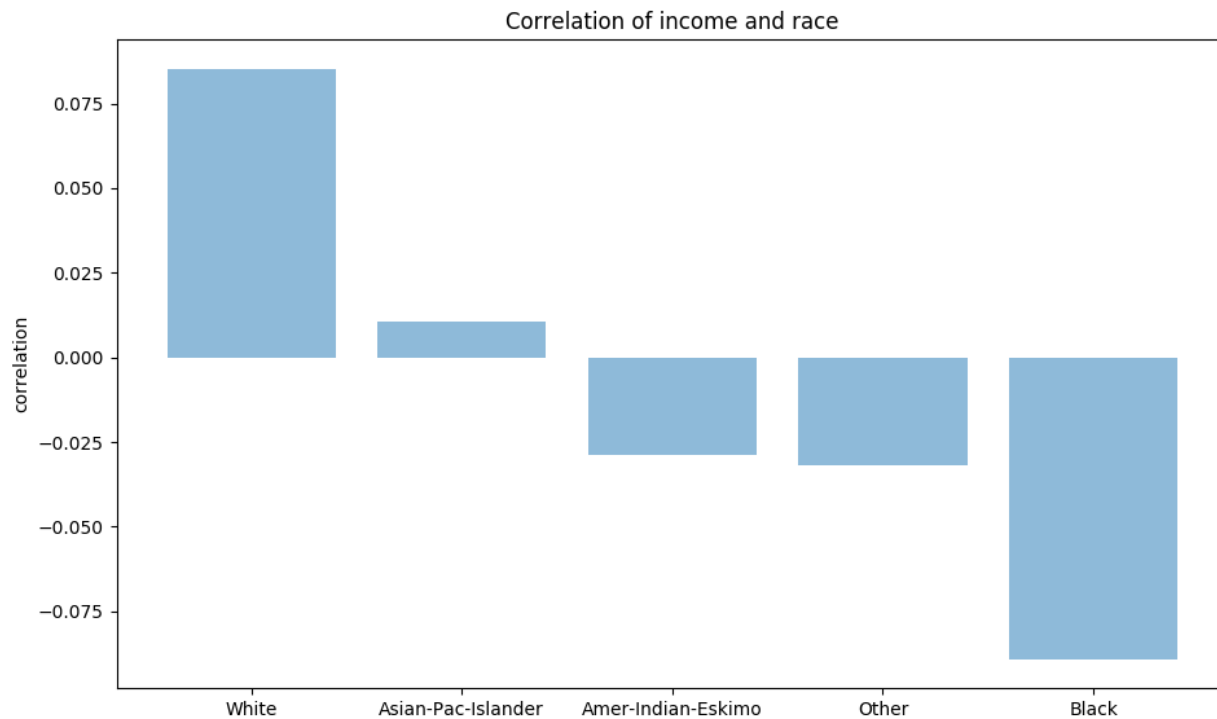


Figure 15 Bar graph of correlation coefficient of race divisions

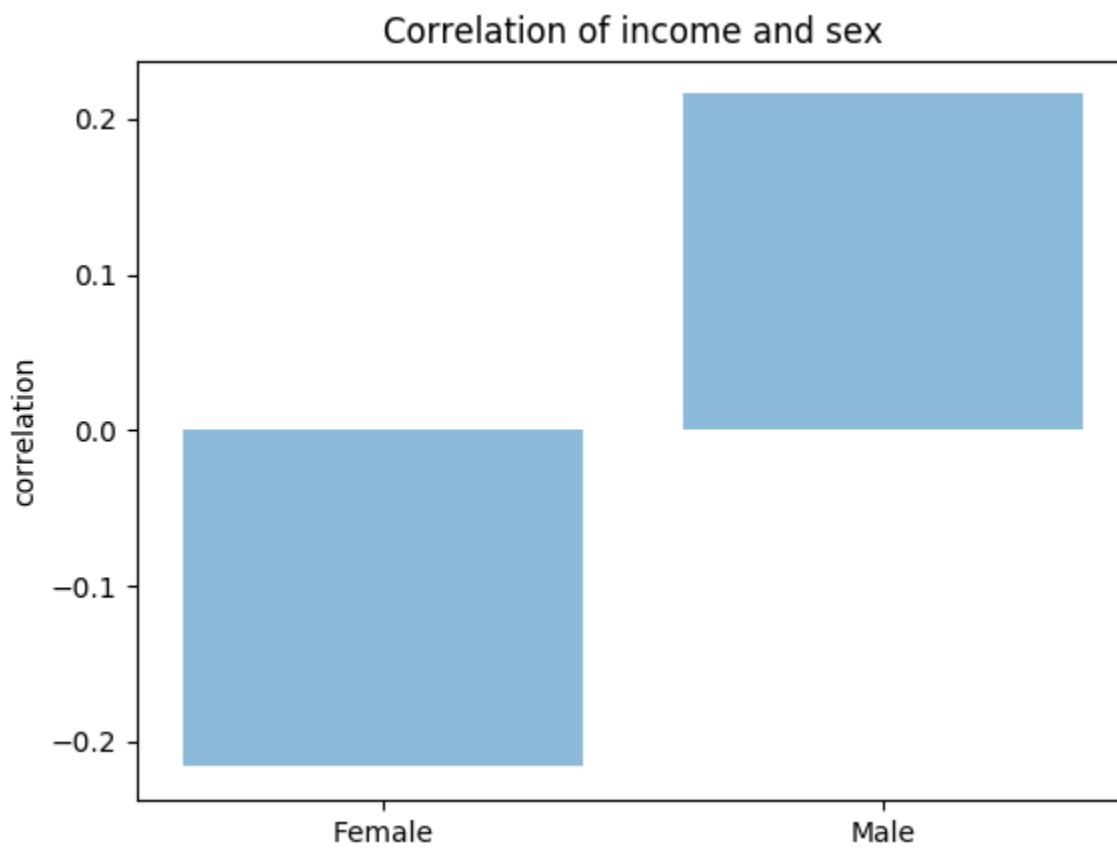


Figure 16 Bar graph of correlation coefficient of sex divisions

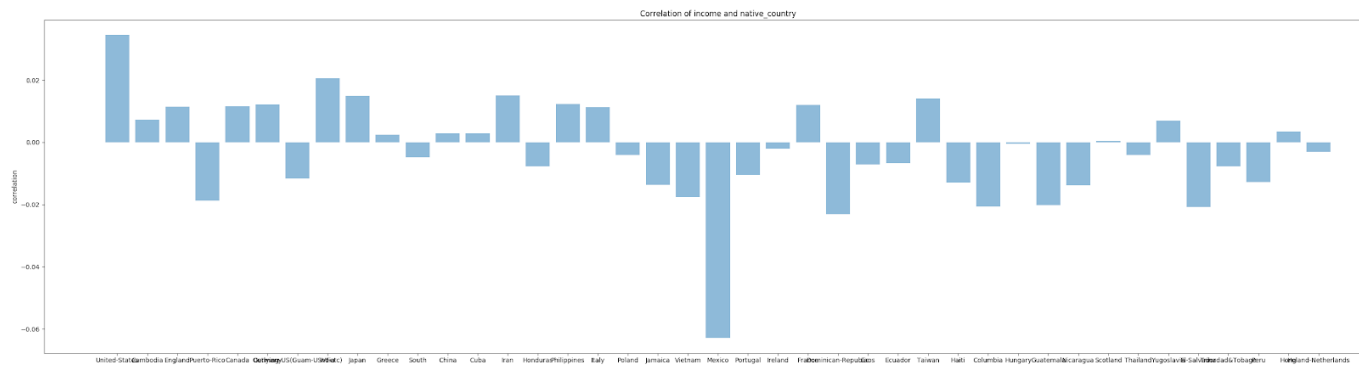


Figure 17 Bar graph of correlation coefficient of native-country divisions

Figure 17 has too many native countries to show properly