- Sex
- Age
- Marital State
- Disability
- Nativity
- Citizenship
- Language Skills
- School Level
- Employment Status

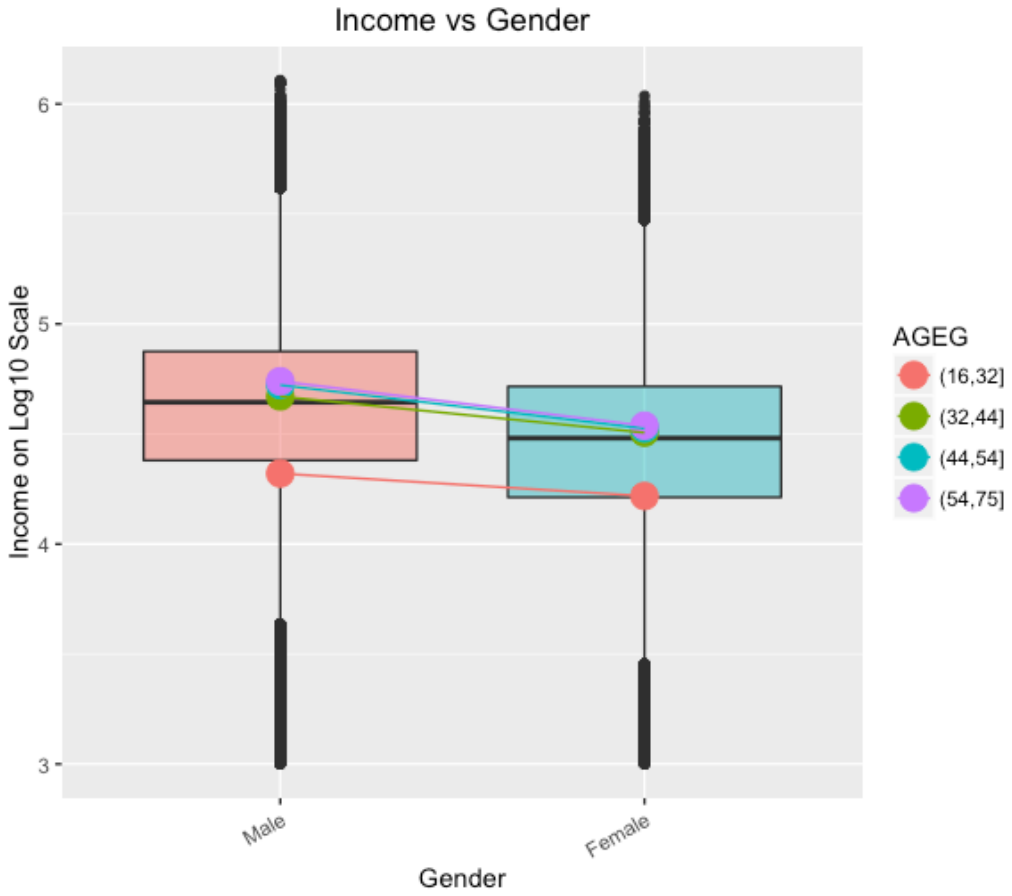
	Estimate	STD Error	P Value
(Intercept)	-0.9116	0.1592	1.018e-08
CITNotCitizen	-0.2039	0.09201	0.02672
AGEP	-0.003014	0.001907	0.1139
MARSingle	-0.5562	0.05486	3.68e-24
SCHLMSc	0.3229	0.05162	3.944e-10
SCHLPhD	0.4372	0.1103	7.392e-05
LANXmultiLang	0.1438	0.06576	0.02875
SEXFEMALE	-0.017	0.04433	0.7013
DISHealthy	1.125	0.1062	3.259e-26
NATIVITYNonNative	-0.08082	0.07654	0.291

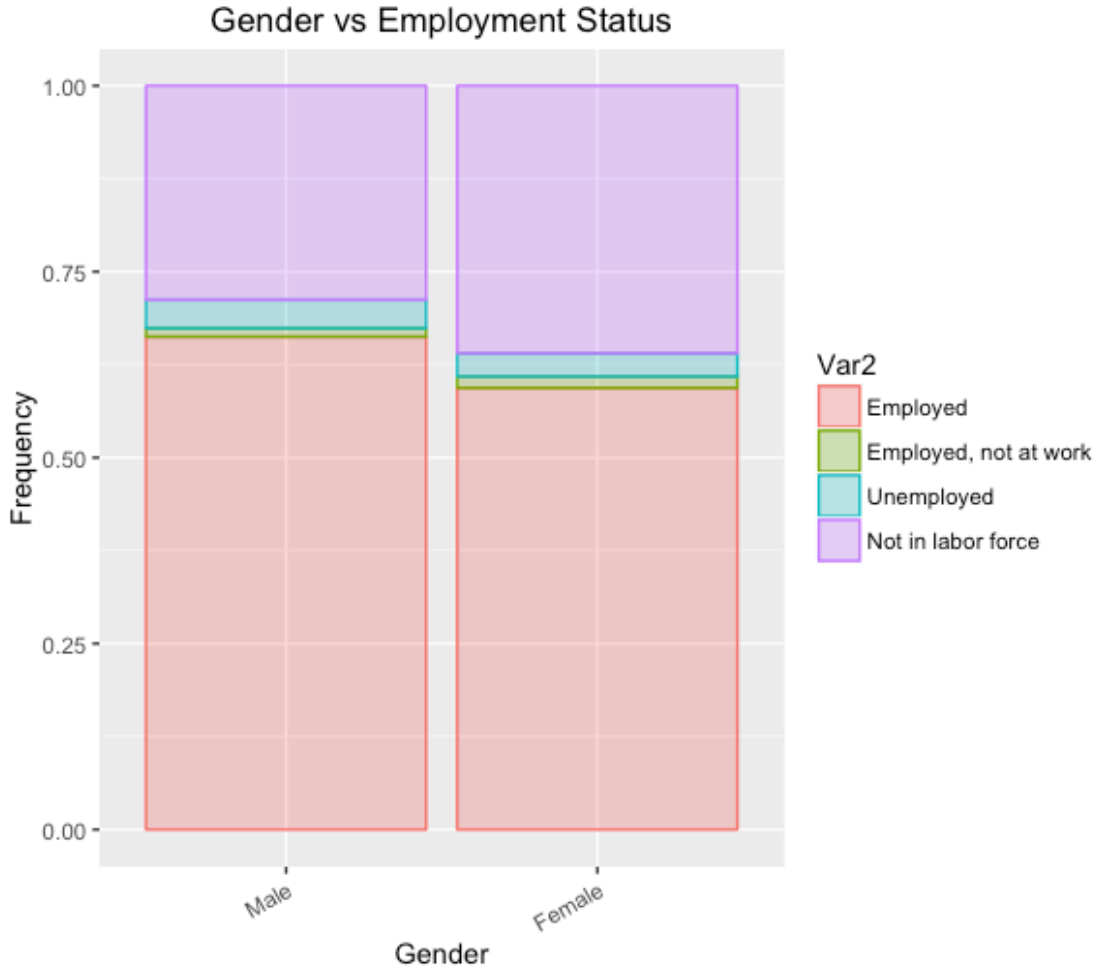
We can see that the correlations of the “Sex”, “Age”, and “Nativity” with the output is not statistically sig
by a collinearity between the input variables, which prevents them from being statistically significant. To
reporting the variance inflation factor between the variables:

	GVIF	Df
CIT	1.439	1
AGEP	1.402	1
MAR	1.329	1
SCHL	1.036	2
LANX	1.886	1
SEX	1.013	1
DIS	1.03	1
NATIVITY	2.305	1

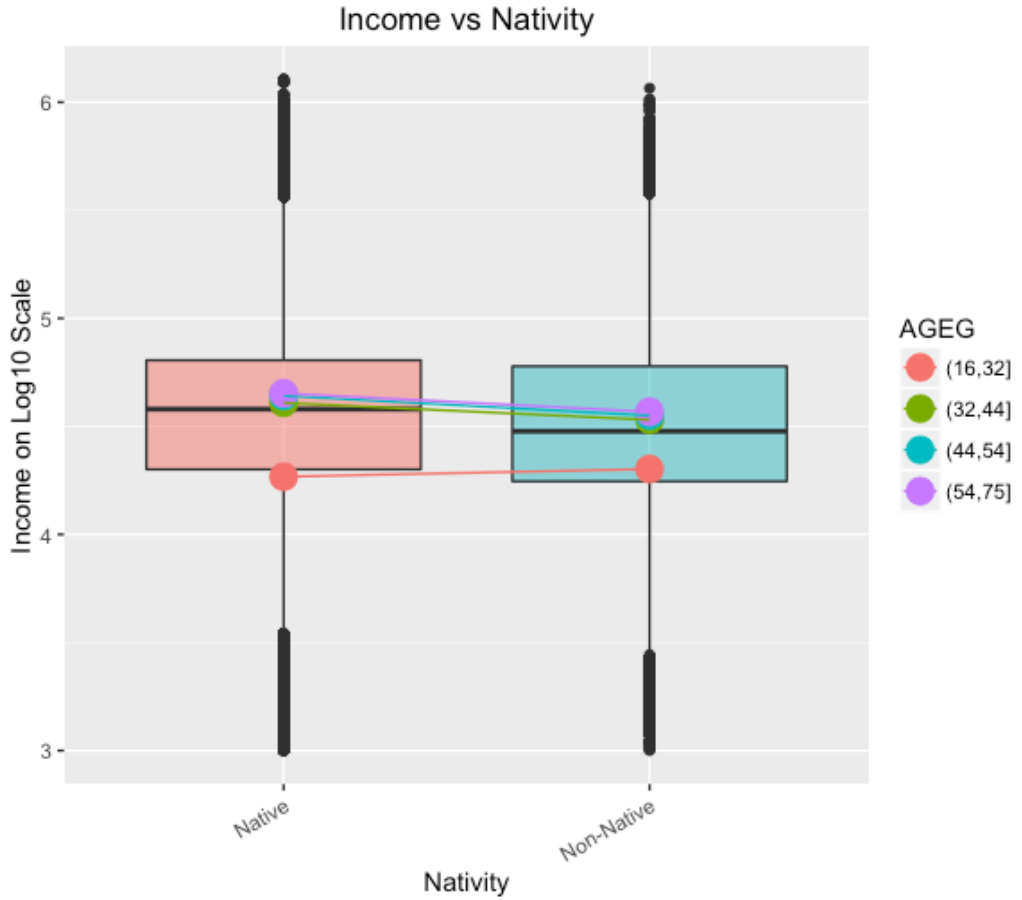
It doesn’t seem like it!

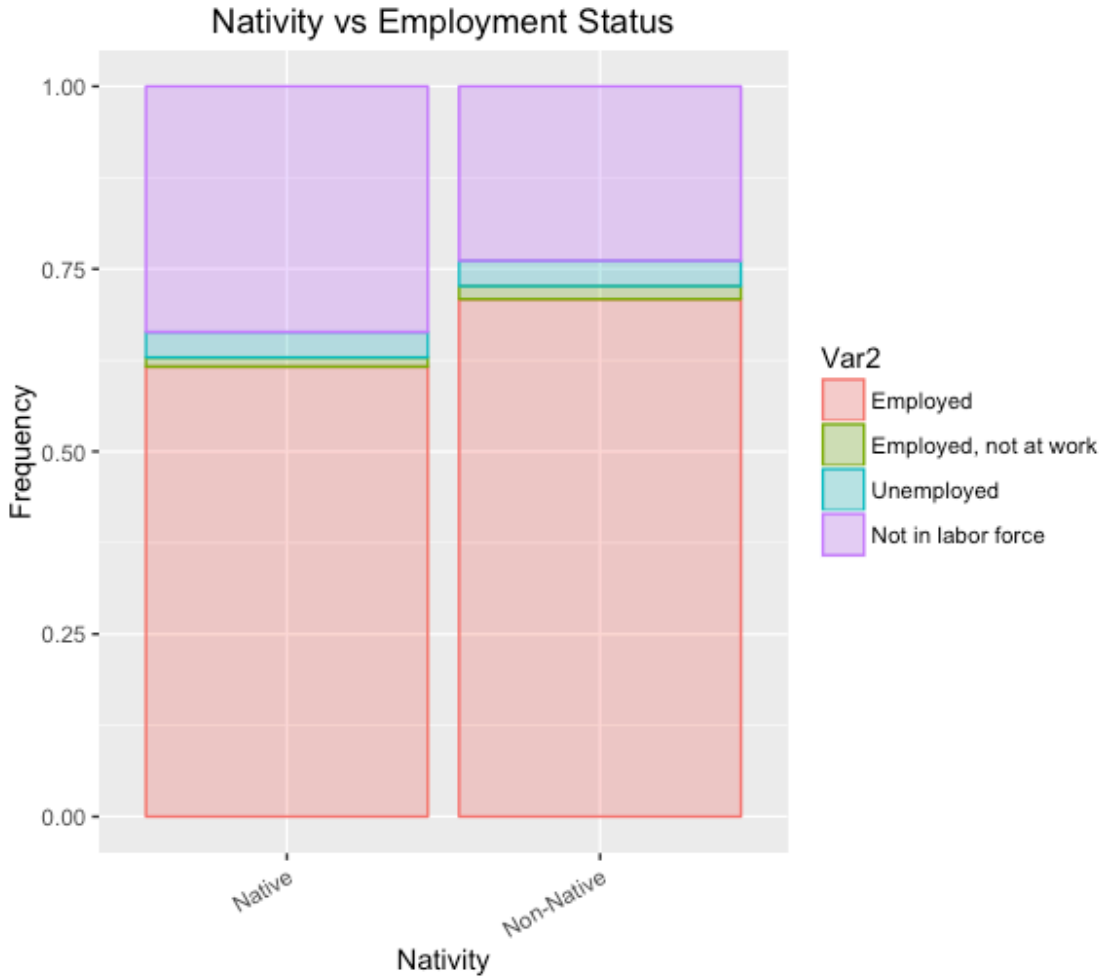
Income vs Sex



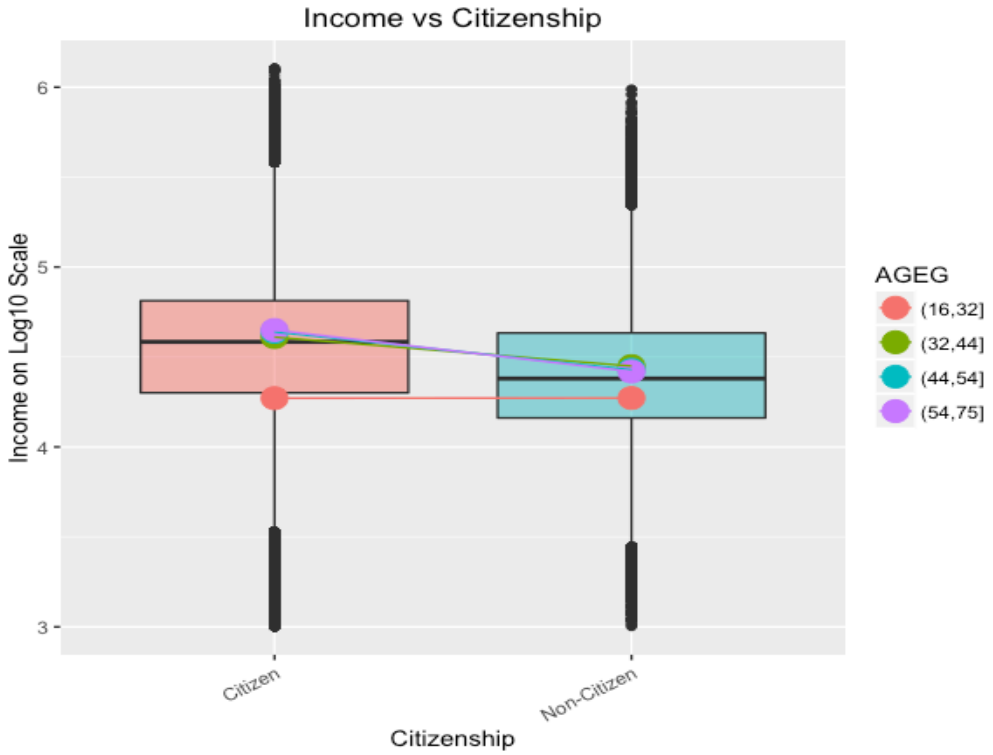


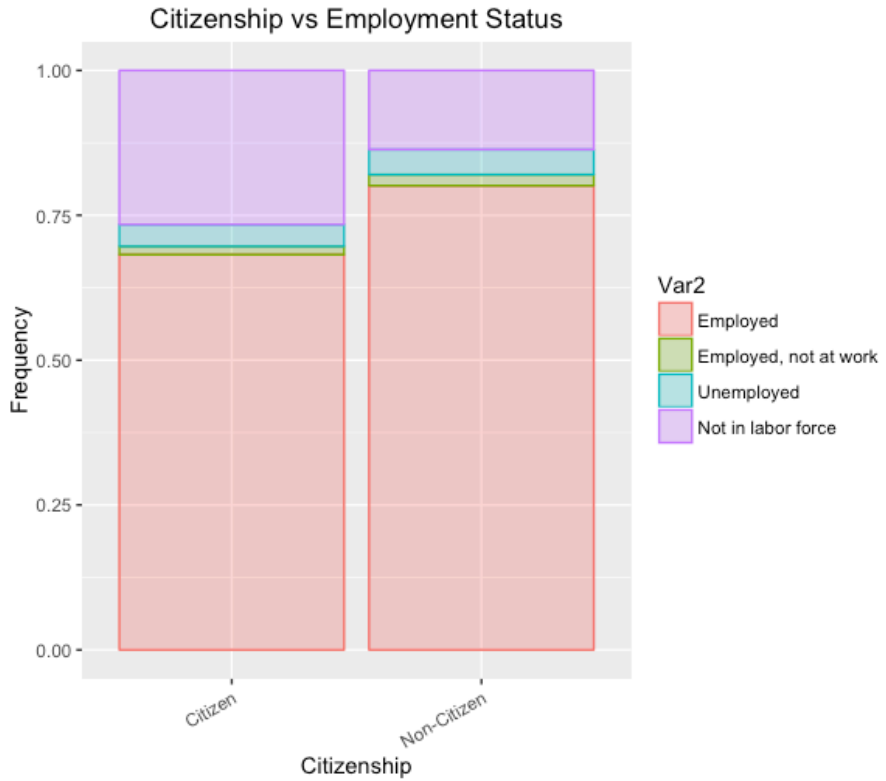
Income vs Nativity



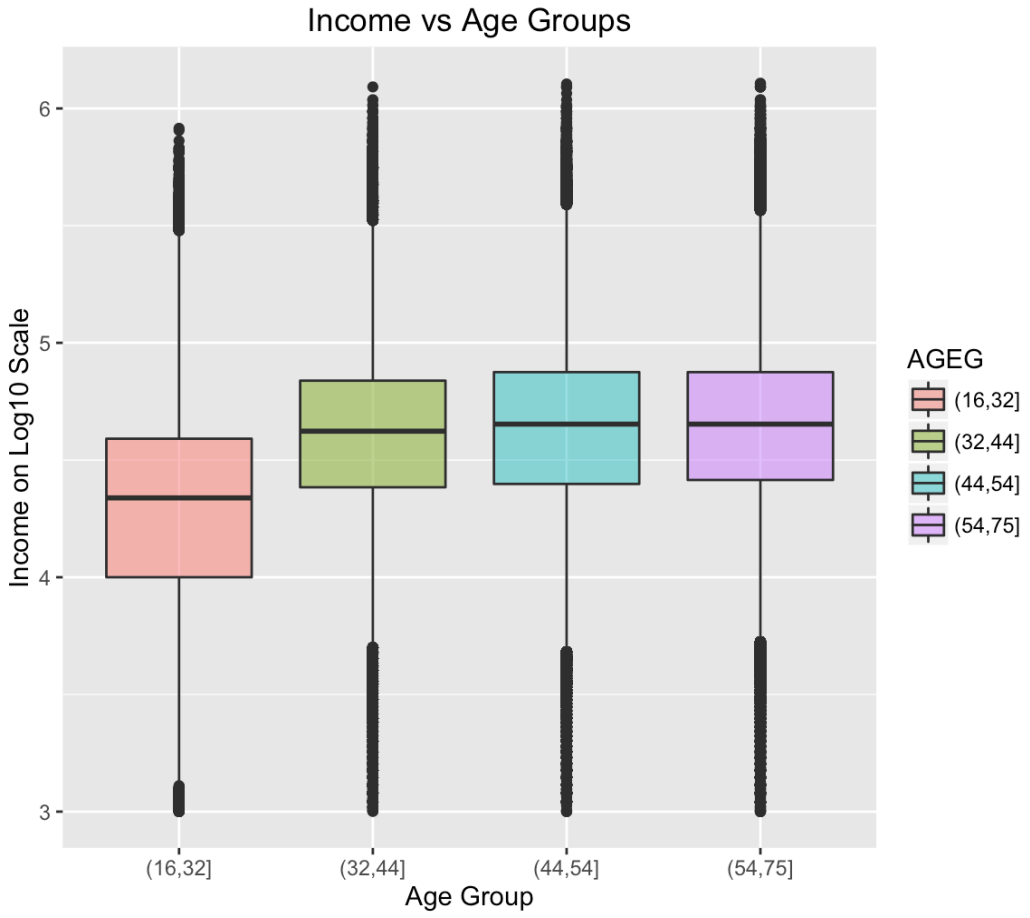


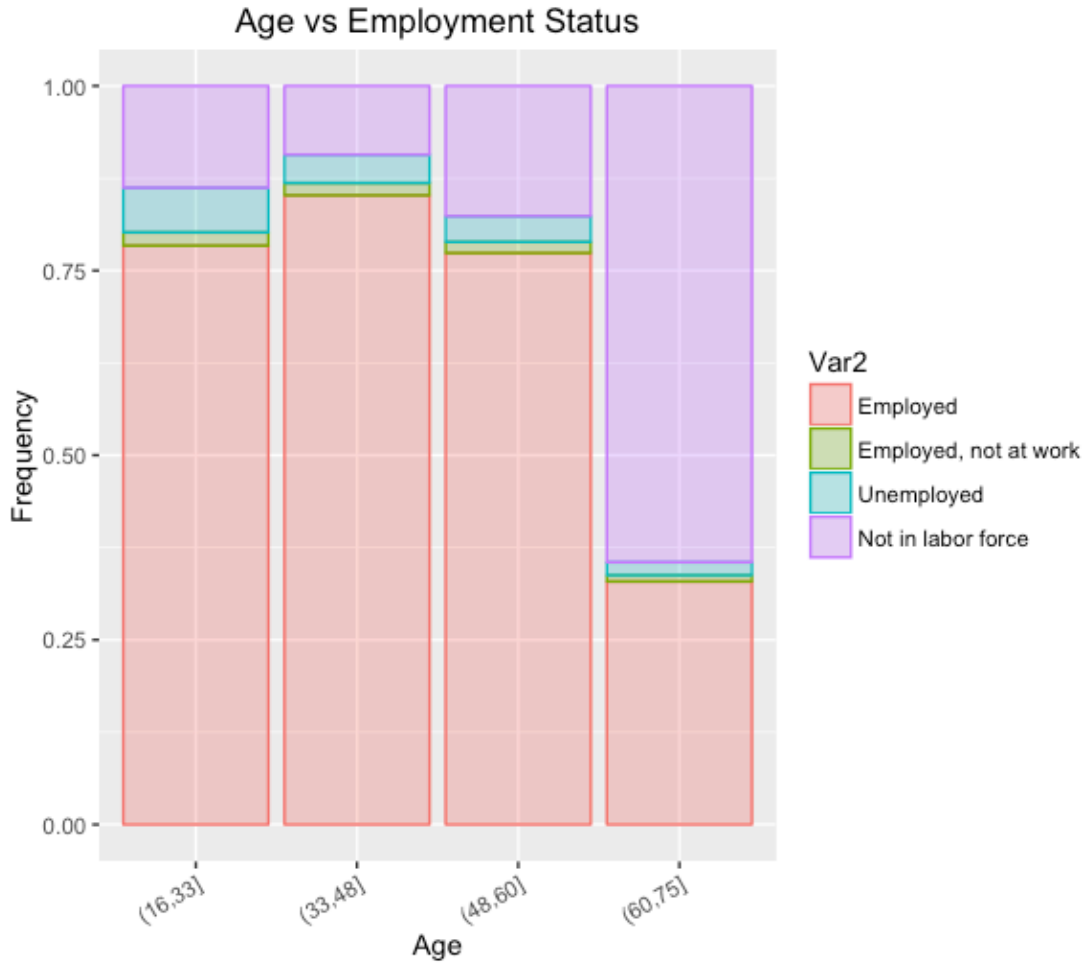
Income vs Citizenship



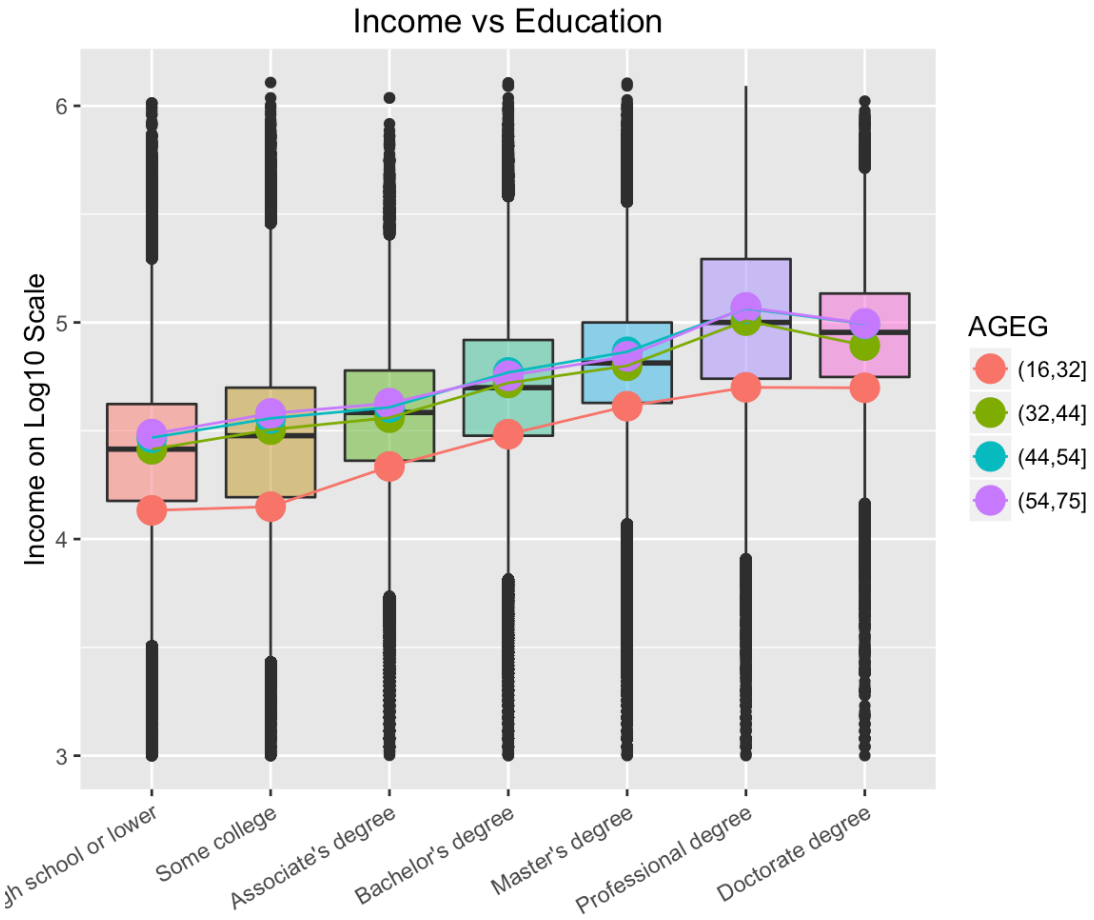


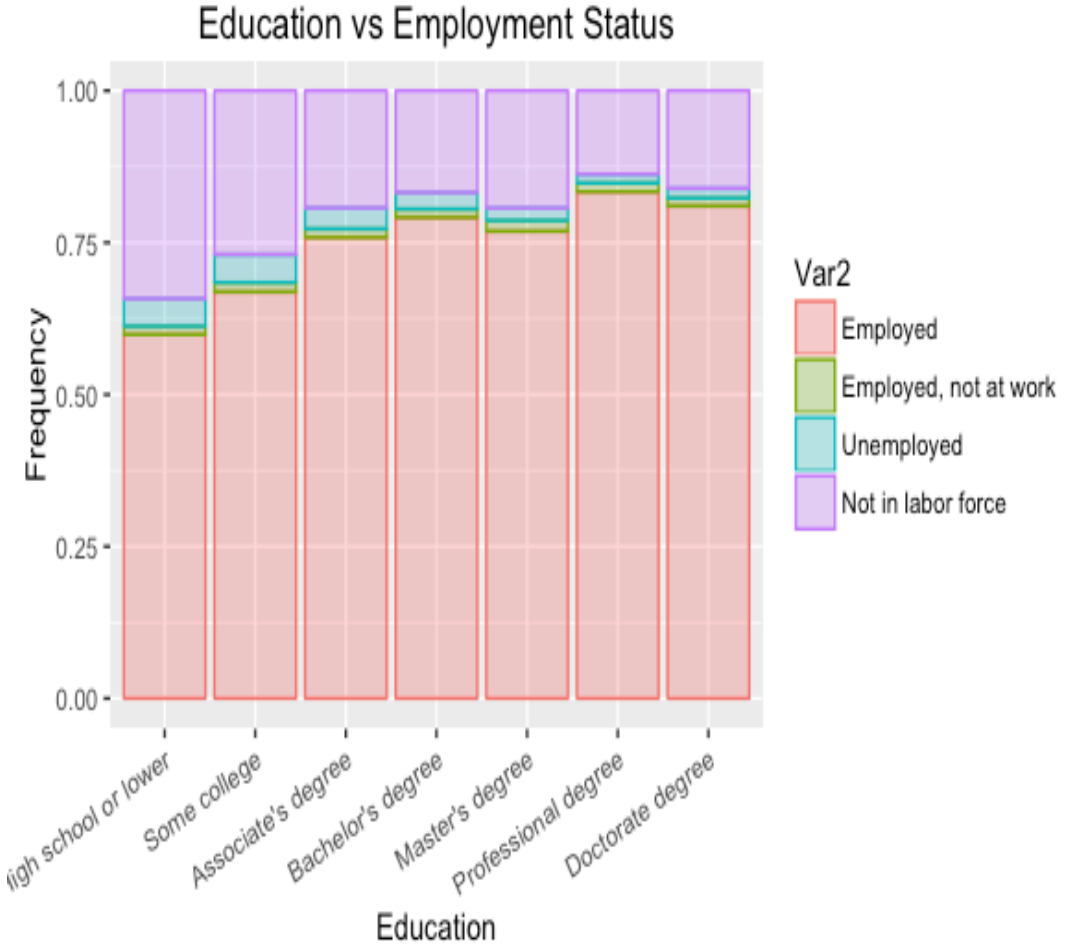
Income vs Age



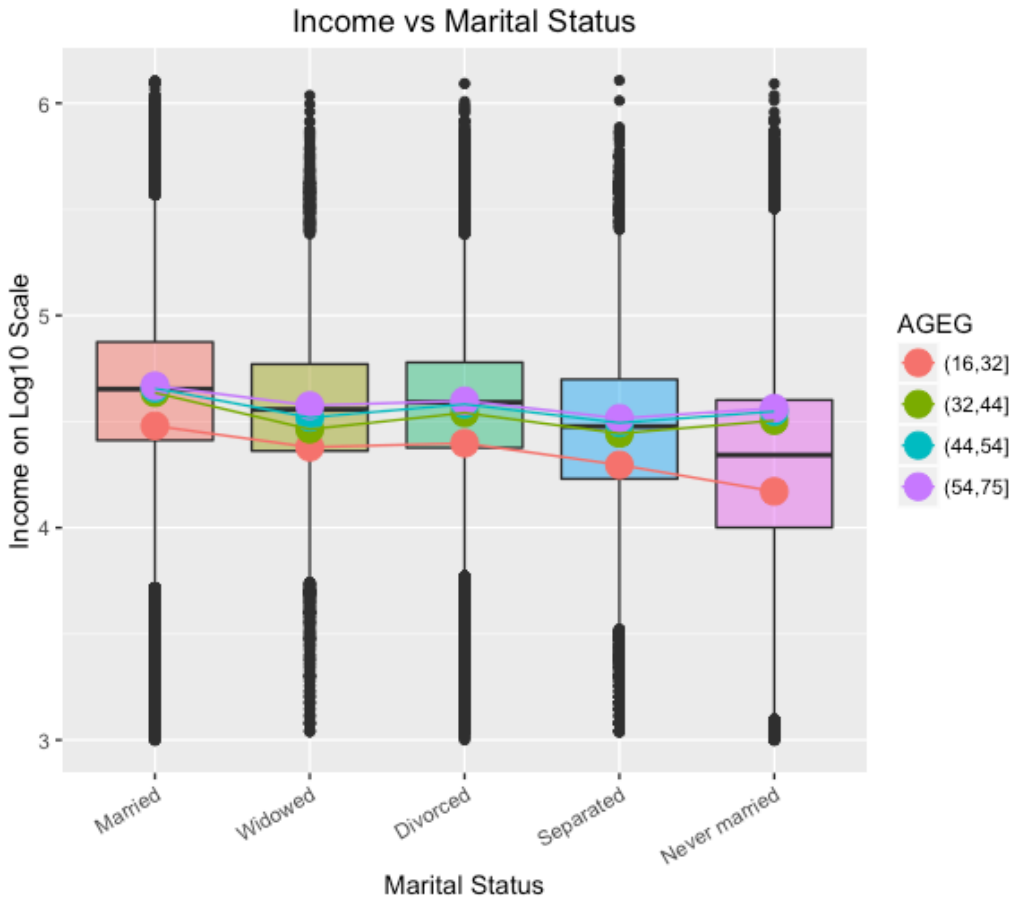


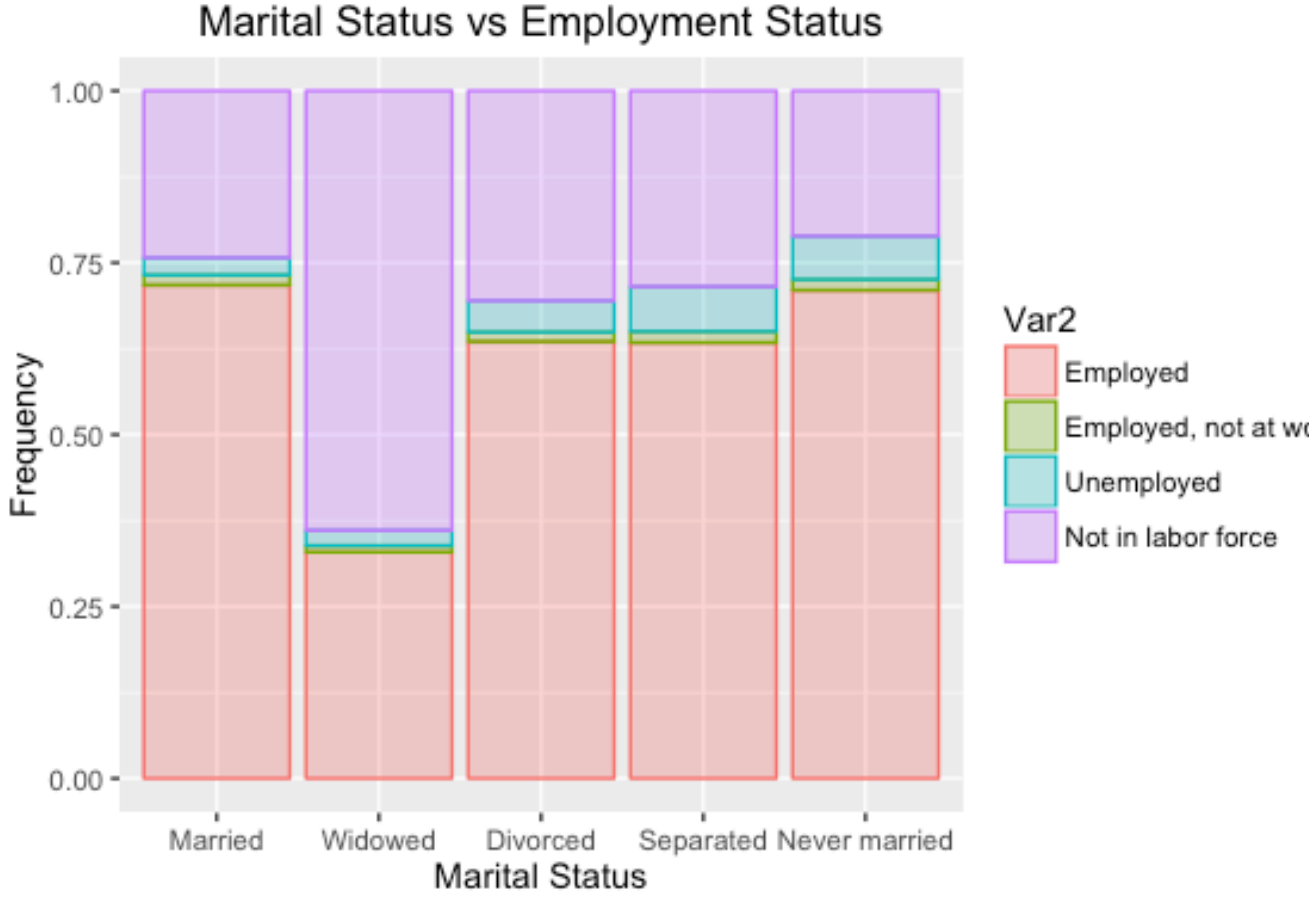
Income vs Education



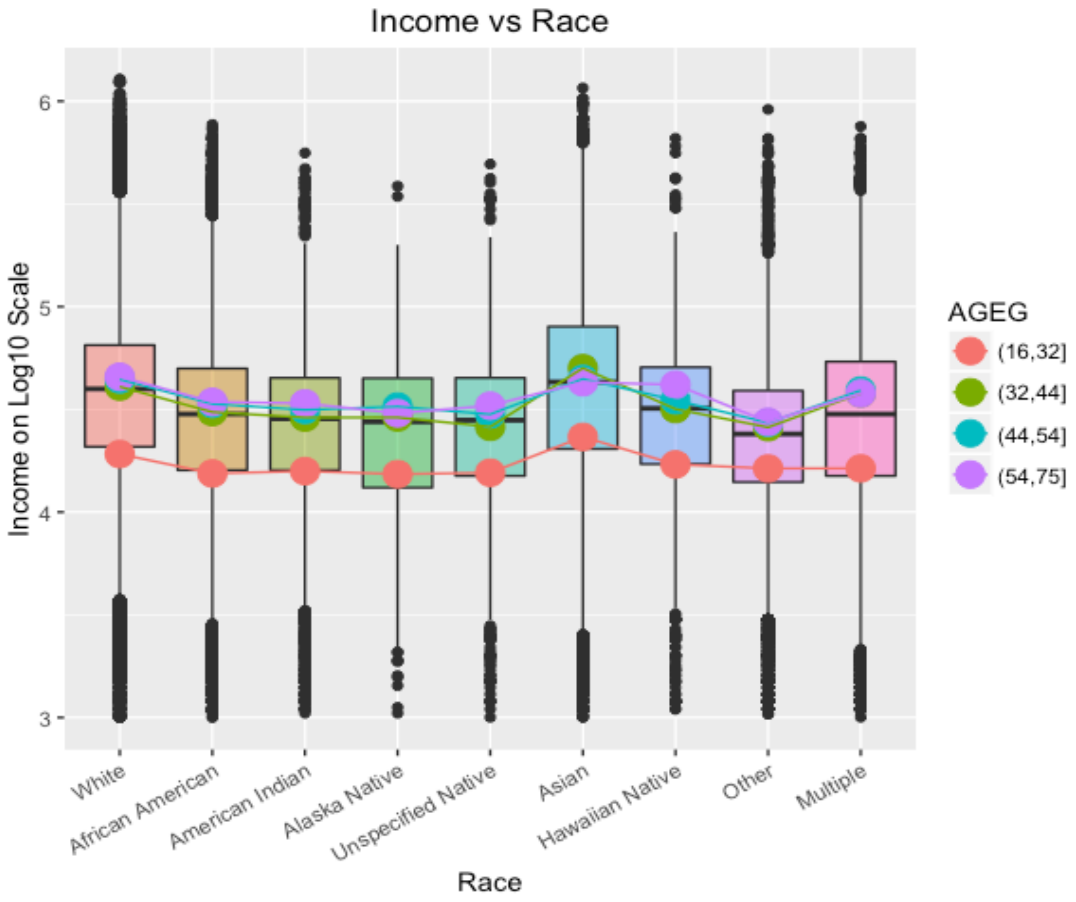


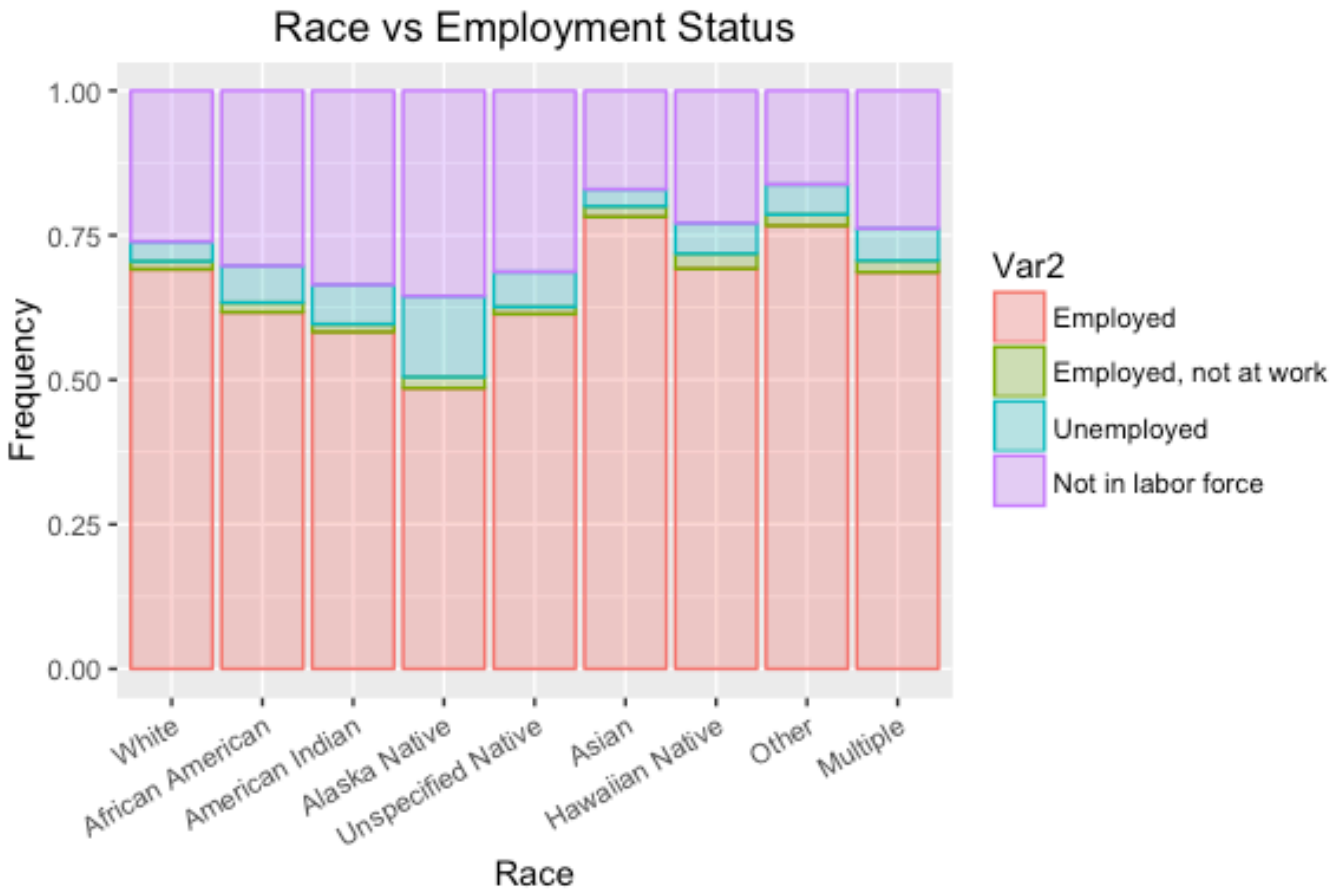
Income vs Marital Status



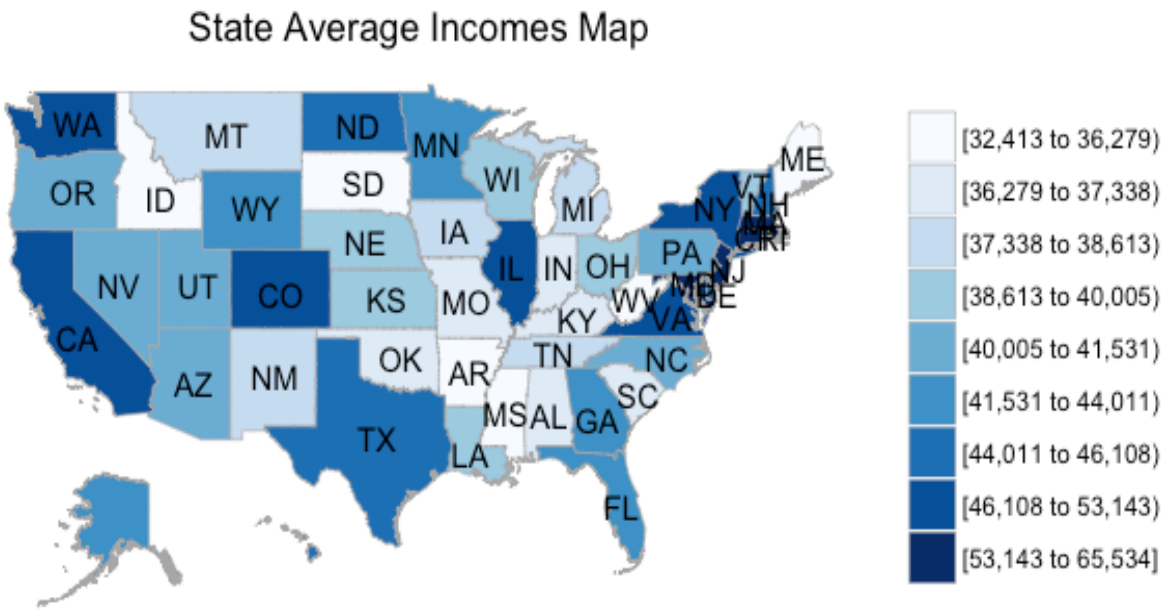


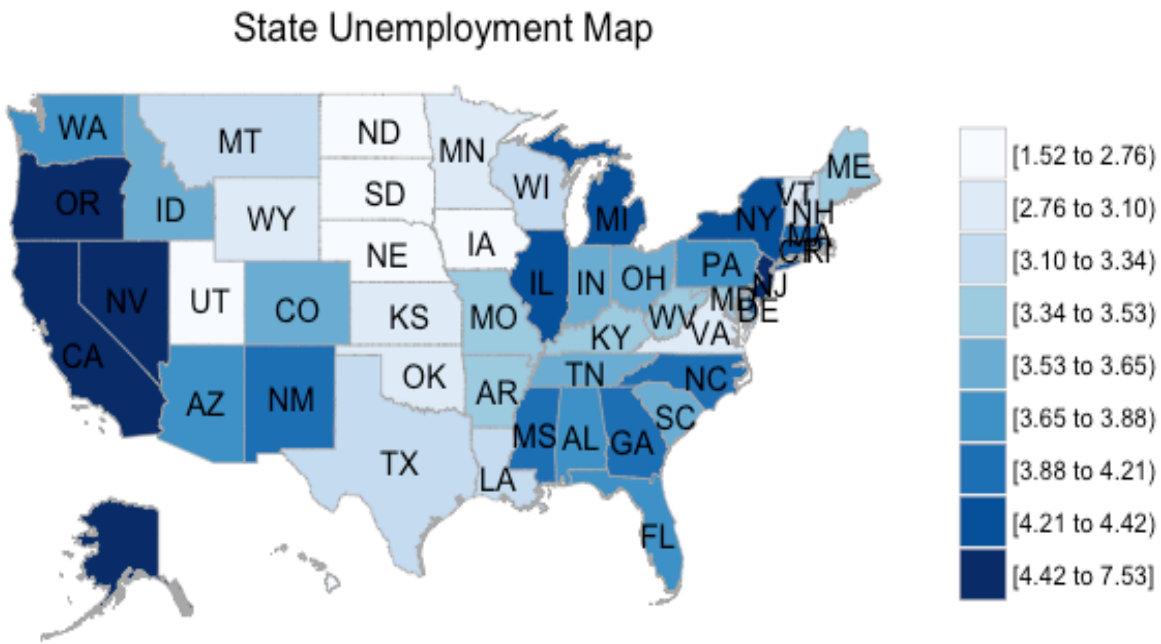
Income vs Race

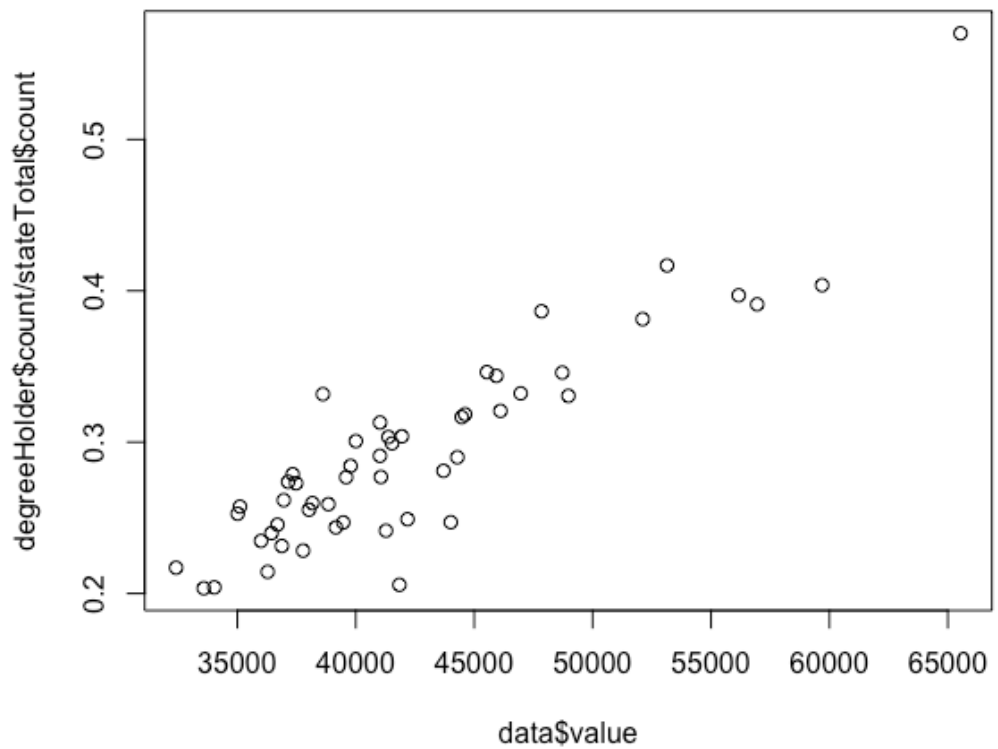




Income vs Employment Map across the States



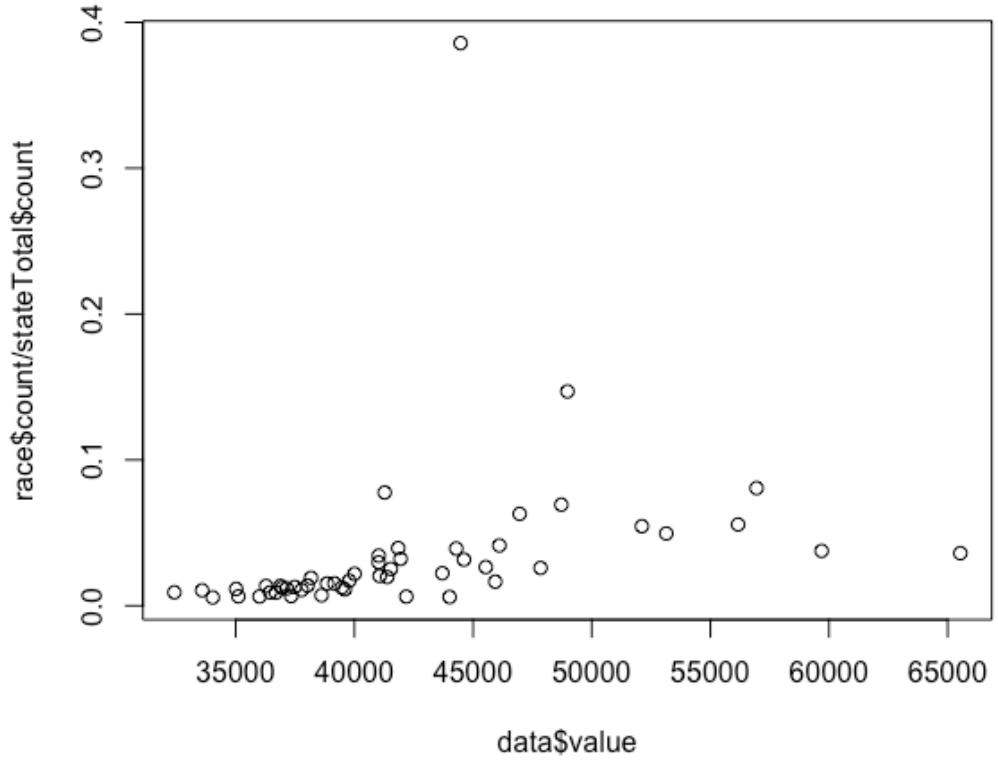



degreeHolder$count/stateTotal$count
0.2
0.3
0.4
0.5
35000
40000
45000
50000
55000
60000
65000
data$value

Cor=0.903

RACE=Asian

Cor=0.33




```
> cor
[1] -0.2078640225  0.2911087046 -0.2412356918  0.0274550146  0.1125843177
[6]  0.3345368383  0.0003149281  0.3367420952  0.1302585273
```

AGE

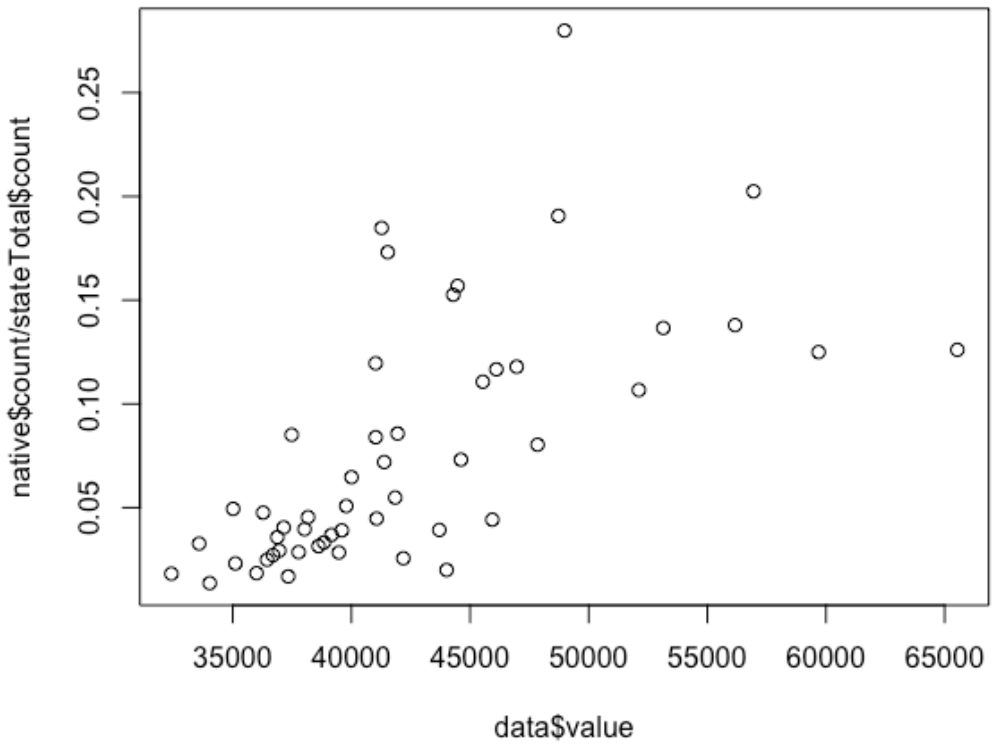
```
> age  <- ds%>%
+   filter(AGEP>=16&AGEP<=32)%>%
+   group_by(ST)%>%
+   summarise(count = n())
> cor(data$value,age$count/stateTotal$count)
[1] 0.4576832
> plot(data$value,age$count/stateTotal$count)
> age  <- ds%>%
+   filter(AGEP>=33&AGEP<=44)%>%
+   group_by(ST)%>%
+   summarise(count = n())
> cor(data$value,age$count/stateTotal$count)
[1] 0.450071
> plot(data$value,age$count/stateTotal$count)
> age  <- ds%>%
+   filter(AGEP>=45&AGEP<=54)%>%
+   group_by(ST)%>%
+   summarise(count = n())
> cor(data$value,age$count/stateTotal$count)
[1] 0.1916612
> plot(data$value,age$count/stateTotal$count)
> age  <- ds%>%
+   filter(AGEP>=55&AGEP<=75)%>%
+   group_by(ST)%>%
+   summarise(count = n())
> cor(data$value,age$count/stateTotal$count)
[1] -0.1506796
```

Marital Status

```
> cor
[1] -0.2036917 -0.5031219 -0.5386787  0.1030570  0.7206929
```

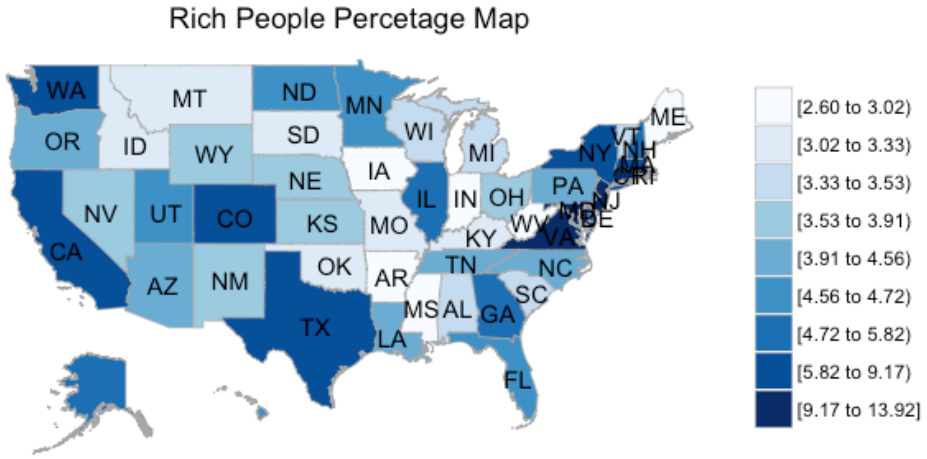
Nativity

Non-nativity vs income
Cor: 0.65



The rich

- The rich: whose annual Income >= $120000



Cor with degreeholder count: 0.91