Wage and Race

Code

Hide

```
# set working directory
setwd('/Users/evali/Library/CloudStorage/OneDrive-UniversityofToronto/Master of Infor
mation/Year 1/Fall 2022/INF 2190 - DATA ANALYTICS/Research Project')
```

Hide

```
# install ISLR2
# install.packages("ISLR2")

# load package
library(ISLR2)
```

Hide

```
# load dataset
data(Wage)
head(Wage)
```

	y...	...	maritl	race	education	region	jobclas
	<int>	<int>	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>
231655	2006	18	1. Never Married	1. White	1. < HS Grad	2. Middle Atlantic	1. Indus
86582	2004	24	1. Never Married	1. White	4. College Grad	2. Middle Atlantic	2. Infor
161300	2003	45	2. Married	1. White	3. Some College	2. Middle Atlantic	1. Indus
155159	2003	43	2. Married	3. Asian	4. College Grad	2. Middle Atlantic	2. Infor
11443	2005	50	4. Divorced	1. White	2. HS Grad	2. Middle Atlantic	2. Infor
376662	2008	54	2. Married	1. White	4. College Grad	2. Middle Atlantic	2. Infor

6 rows | 1-8 of 11 columns

Data Cleaning and Pre-processing

Hide

```
# checking for missing data
is.null(Wage)
```

```
[1] FALSE
```

Hide

```
# structure of the data
str(Wage)
```

```
'data.frame':   3000 obs. of  11 variables:
 $ year      : int  2006 2004 2003 2003 2005 2008 2009 2008 2006 2004 ...
 $ age       : int  18 24 45 43 50 54 44 30 41 52 ...
 $ maritl    : Factor w/ 5 levels "1. Never Married",..: 1 1 2 2 4 2 2 1 1 2 ...
 $ race      : Factor w/ 4 levels "1. White","2. Black",..: 1 1 1 3 1 1 4 3 2 1 ...
 $ education : Factor w/ 5 levels "1. < HS Grad",..: 1 4 3 4 2 4 3 3 3 2 ...
 $ region    : Factor w/ 9 levels "1. New England",..: 2 2 2 2 2 2 2 2 2 2 ...
 $ jobclass  : Factor w/ 2 levels "1. Industrial",..: 1 2 1 2 2 2 1 2 2 2 ...
 $ health    : Factor w/ 2 levels "1. <=Good","2. >=Very Good": 1 2 1 2 1 2 2 1 2 2 .
..
 $ health_ins: Factor w/ 2 levels "1. Yes","2. No": 2 2 1 1 1 1 1 1 1 1 ...
 $ logwage   : num  4.32 4.26 4.88 5.04 4.32 ...
 $ wage      : num  75 70.5 131 154.7 75 ...
```

Hide

```
# Dropping region - ALL DATA TAKEN FROM MID-ATLANTIC
data <- subset(Wage, select = -c(region))
head(data)
```

	y...	...	maritl	race	education	jobclass	health
	<int>	<int>	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>
231655	2006	18	1. Never Married	1. White	1. < HS Grad	1. Industrial	1. <=Good
86582	2004	24	1. Never Married	1. White	4. College Grad	2. Information	2. >=Very Go
161300	2003	45	2. Married	1. White	3. Some College	1. Industrial	1. <=Good
155159	2003	43	2. Married	3. Asian	4. College Grad	2. Information	2. >=Very Go
11443	2005	50	4. Divorced	1. White	2. HS Grad	2. Information	1. <=Good
376662	2008	54	2. Married	1. White	4. College Grad	2. Information	2. >=Very Go

6 rows | 1-8 of 10 columns

Part 1: Exploratory Data Analysis - OVERALL

Summary Statistics

Hide

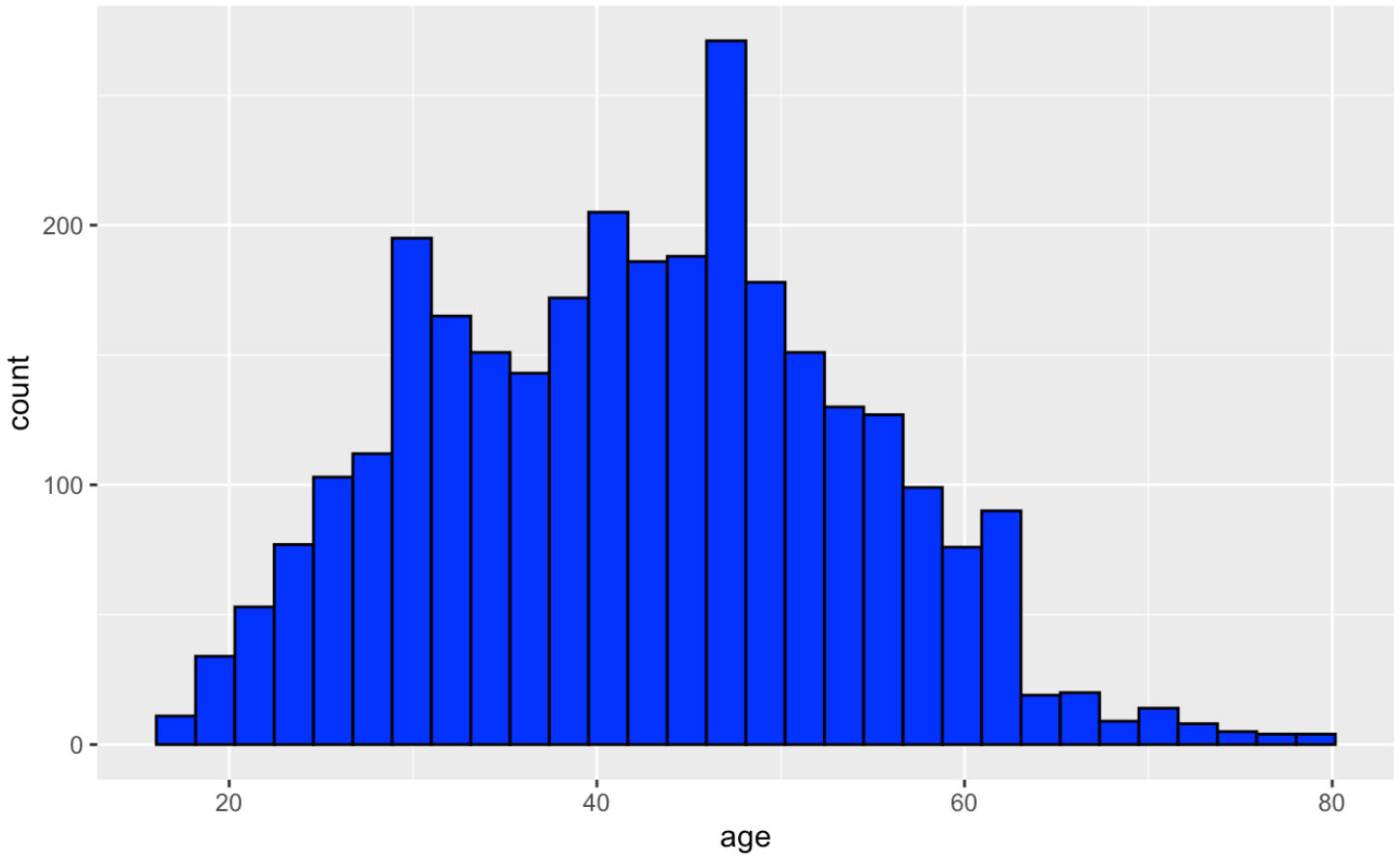
```
summary(data)
```

```
      year           age                     maritl           race
education             jobclass               health
 Min.   :2003   Min.   :18.00   1. Never Married: 648   1. White:2480   1. < HS Grad
:268   1. Industrial :1544   1. <=Good     : 858
 1st Qu.:2004   1st Qu.:33.75   2. Married      :2074   2. Black: 293   2. HS Grad
:971   2. Information:1456   2. >=Very Good:2142
 Median :2006   Median :42.00   3. Widowed      :  19   3. Asian: 190   3. Some Colle
ge    :650
 Mean   :2006   Mean   :42.41   4. Divorced     : 204   4. Other:  37   4. College Gr
ad    :685
 3rd Qu.:2008   3rd Qu.:51.00   5. Separated    :  55                   5. Advanced D
egree:426
 Max.   :2009   Max.   :80.00
  health_ins      logwage           wage
 1. Yes:2083   Min.   :3.000   Min.   : 20.09
 2. No : 917   1st Qu.:4.447   1st Qu.: 85.38
               Median :4.653   Median :104.92
               Mean   :4.654   Mean   :111.70
               3rd Qu.:4.857   3rd Qu.:128.68
               Max.   :5.763   Max.   :318.34
```

Distribution about the dataset

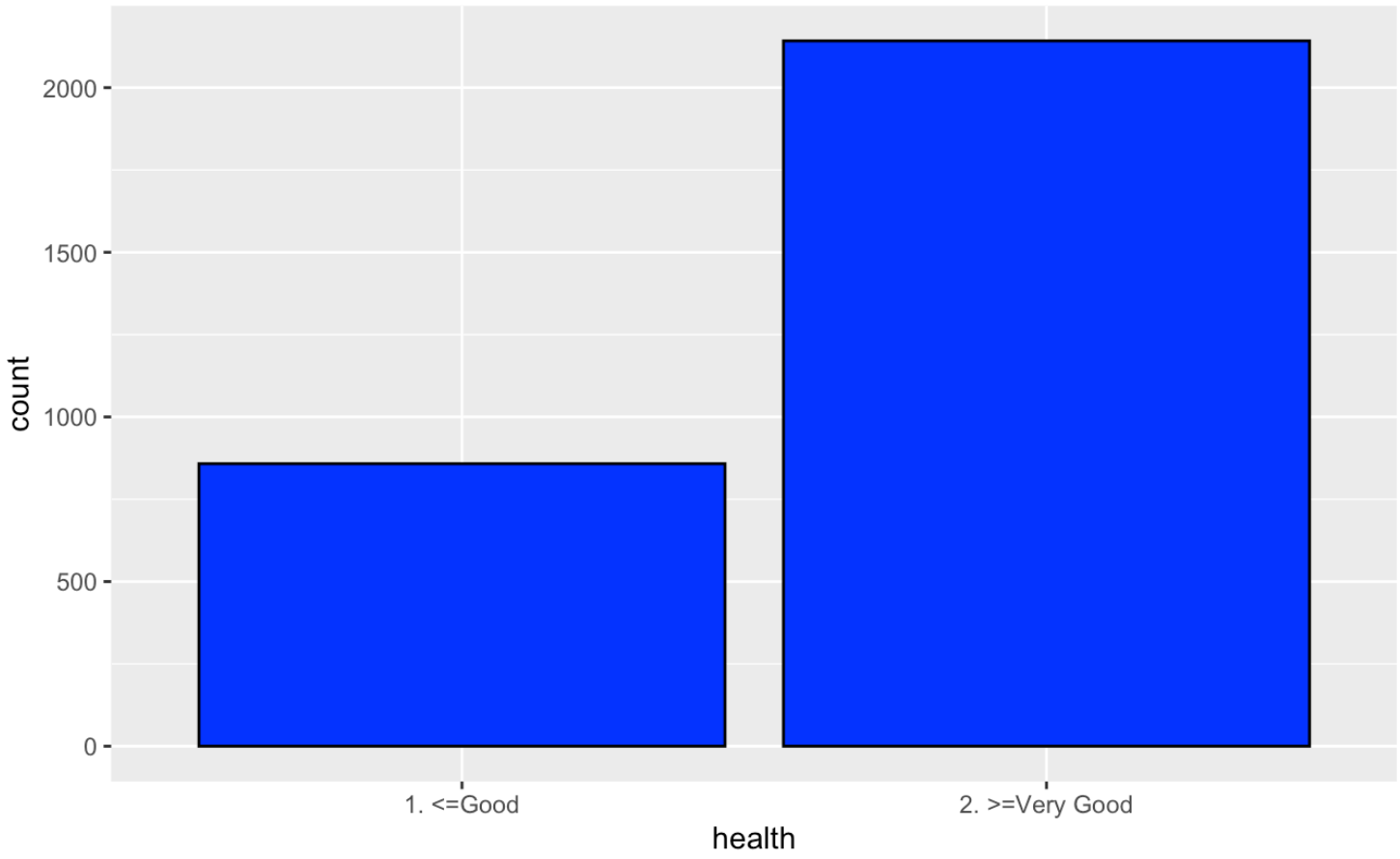
Hide

```
# distribution of age
ggplot(data=data, aes(x=age)) +
    geom_histogram(color='black', fill='blue')
```



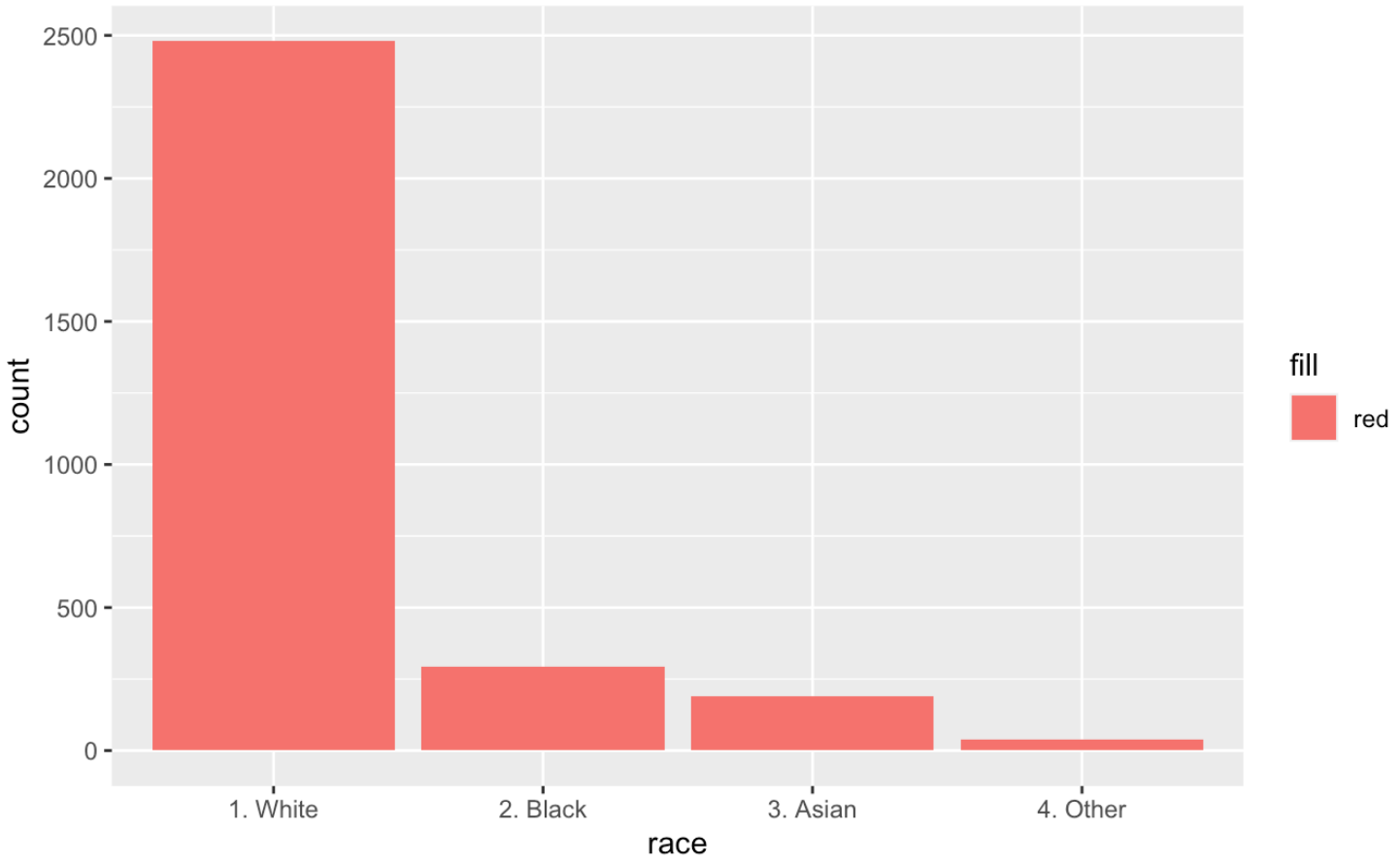
Hide

```
# distribution of edcuation
ggplot(data=data) +
    geom_bar(aes(health), color='black', fill='blue')
```



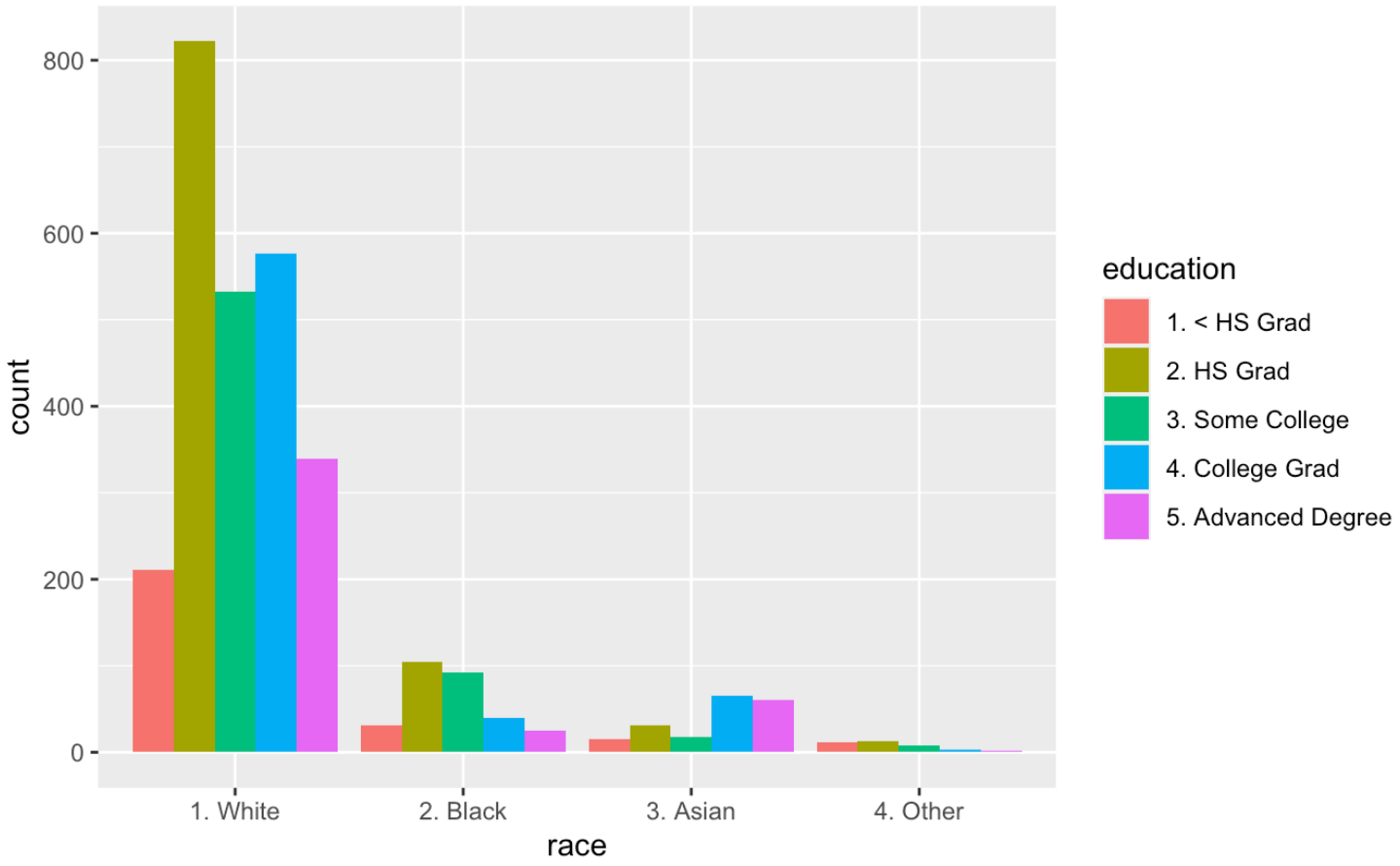
Hide

```
# wage distribution by race
ggplot(data=data) +
  geom_bar(aes(race, fill="red"))
```



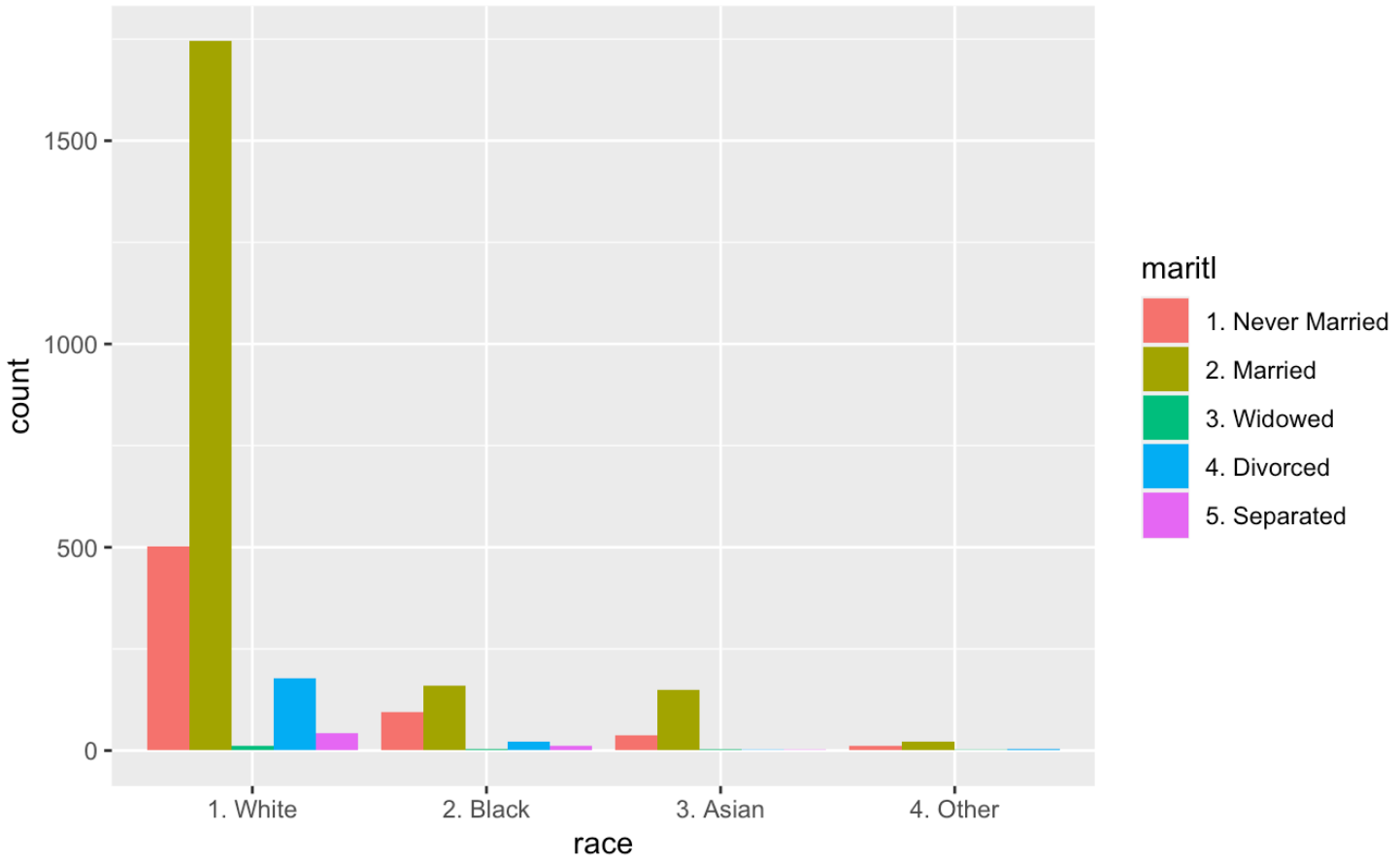
Hide

```
# Distribution of education by race
ggplot(data, aes(race, fill=education )) +
    geom_bar(position="dodge")
```



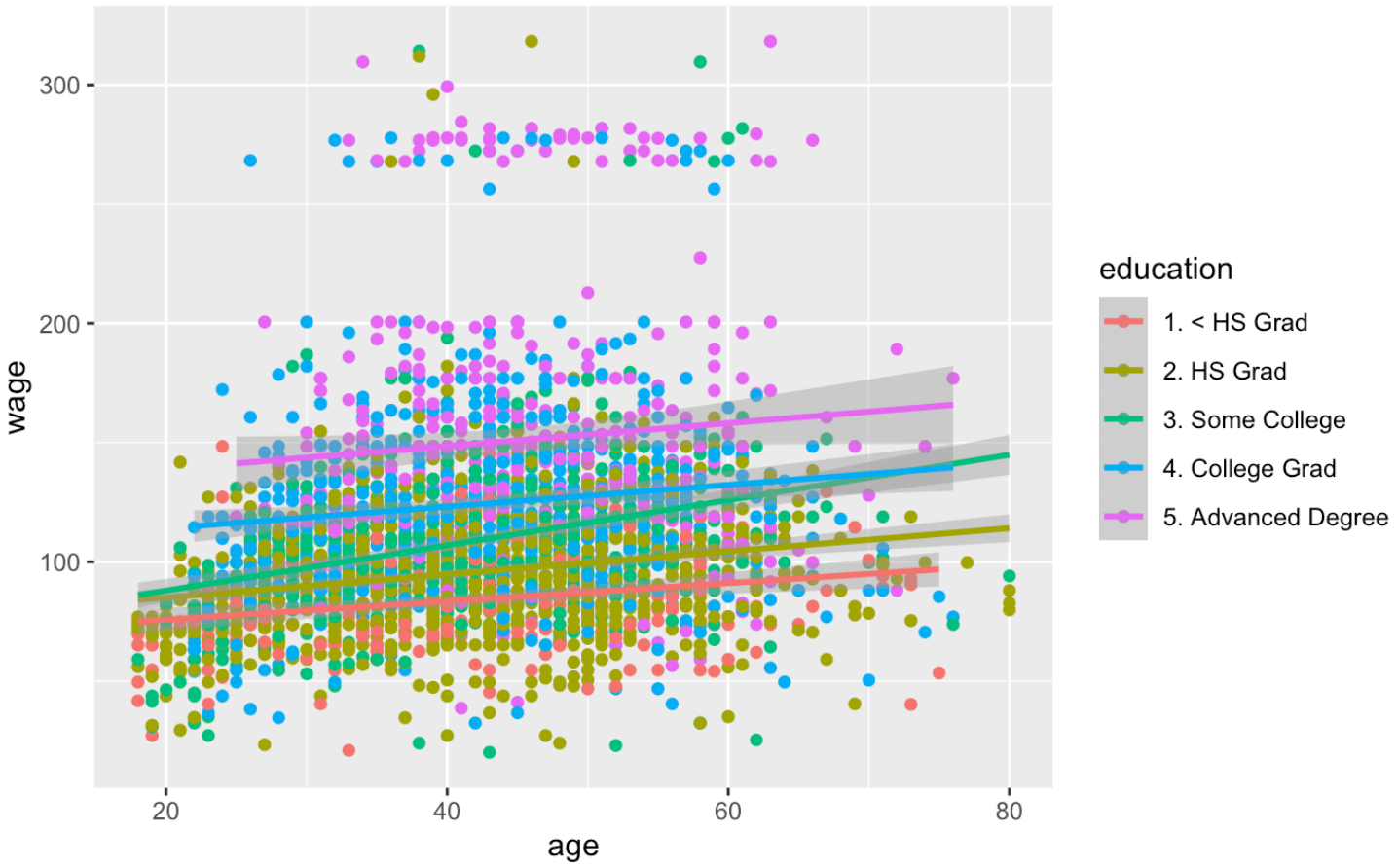
Hide

```
# Distribution of education by race
ggplot(data, aes(race, fill=maritl )) +
    geom_bar(position="dodge")
```



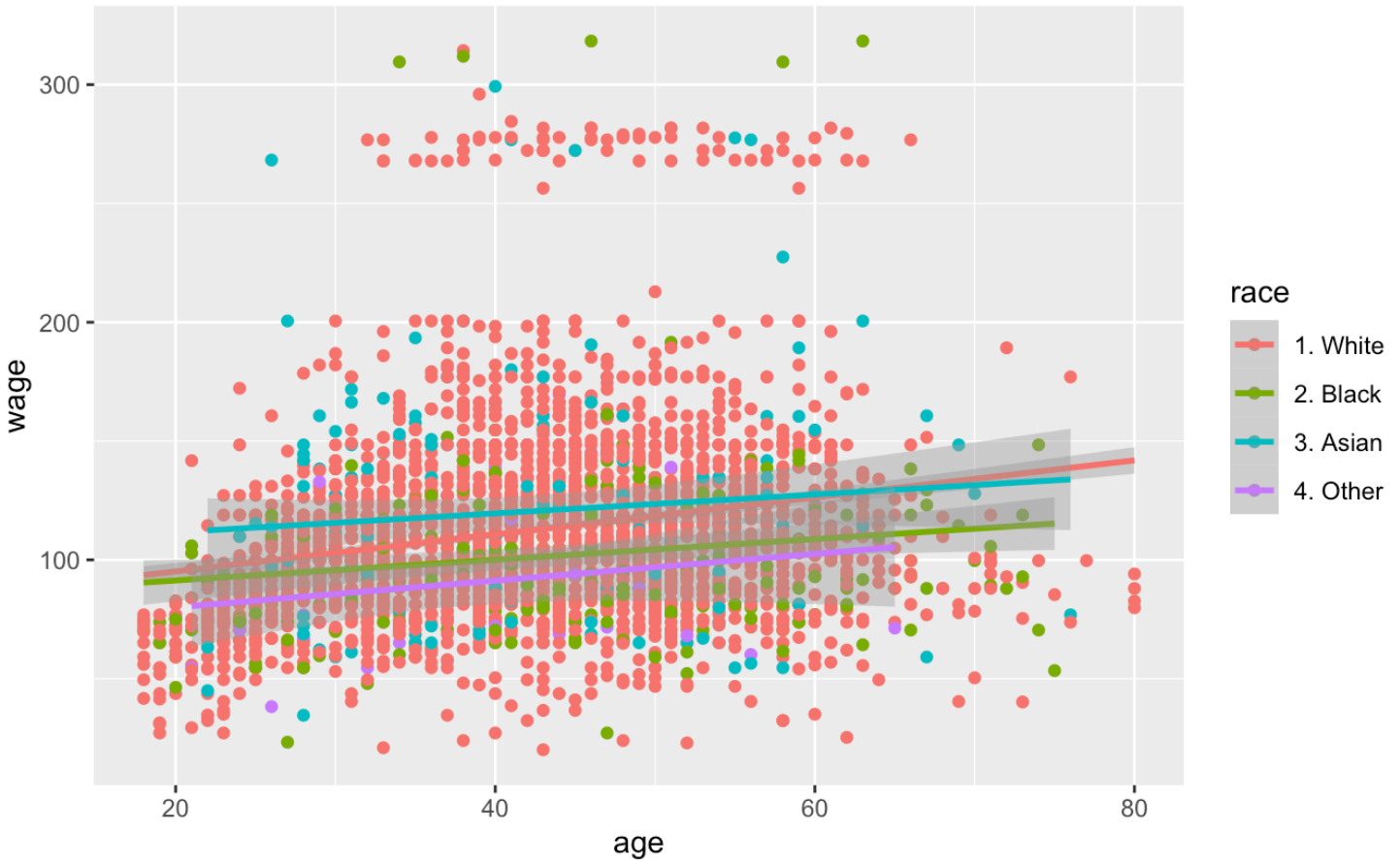
Hide

```
# Wage and age by education
ggplot(data=data, aes(age, wage, color=education)) +
  geom_point() +
  geom_smooth(method="lm", formula = y ~ x)
```



Hide

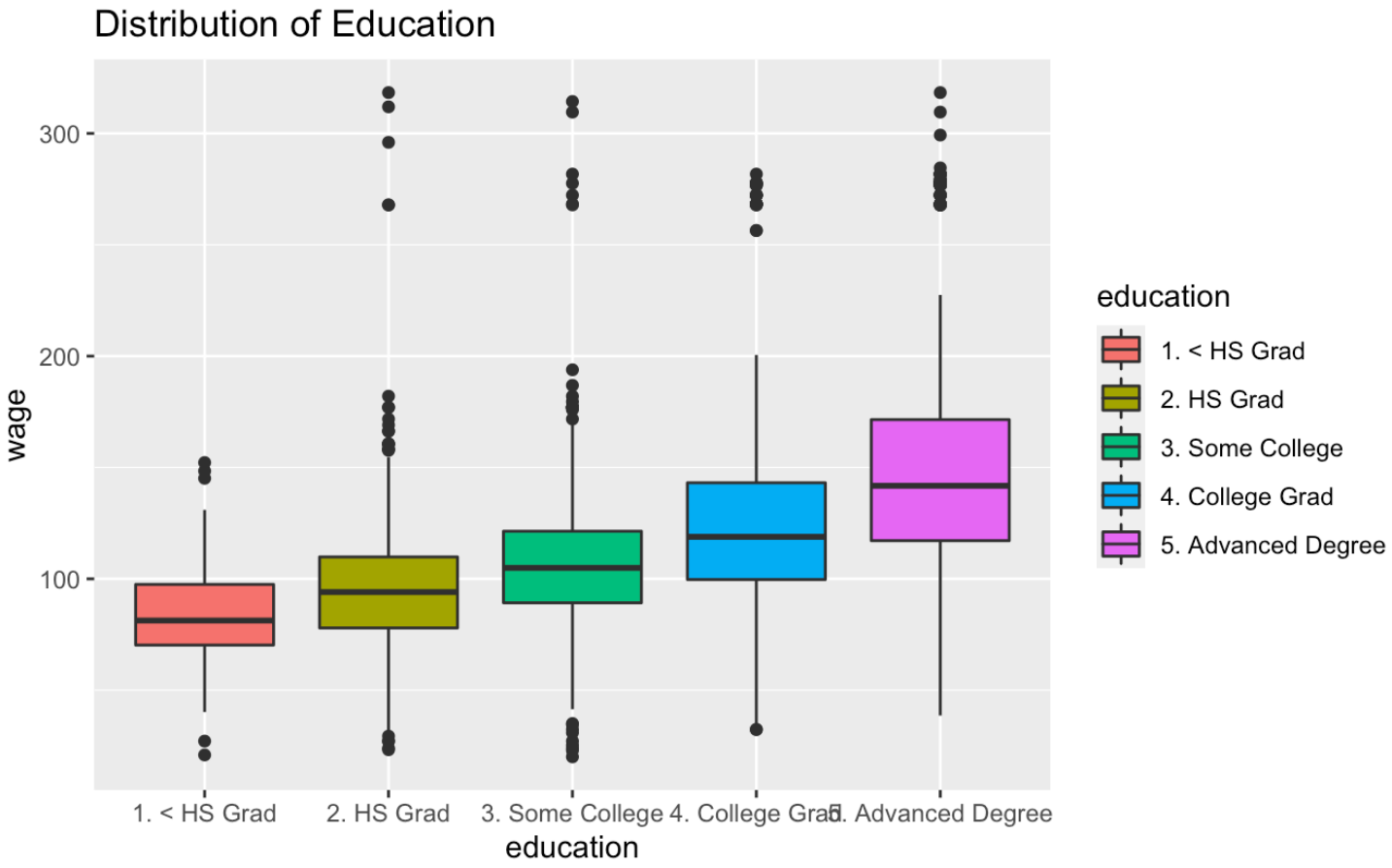
```
# Wage and age by race
ggplot(data=data, aes(age, wage, color=race)) +
  geom_point() +
  geom_smooth(method="lm", formula = y ~ x)
```



Wage Distribution by Education

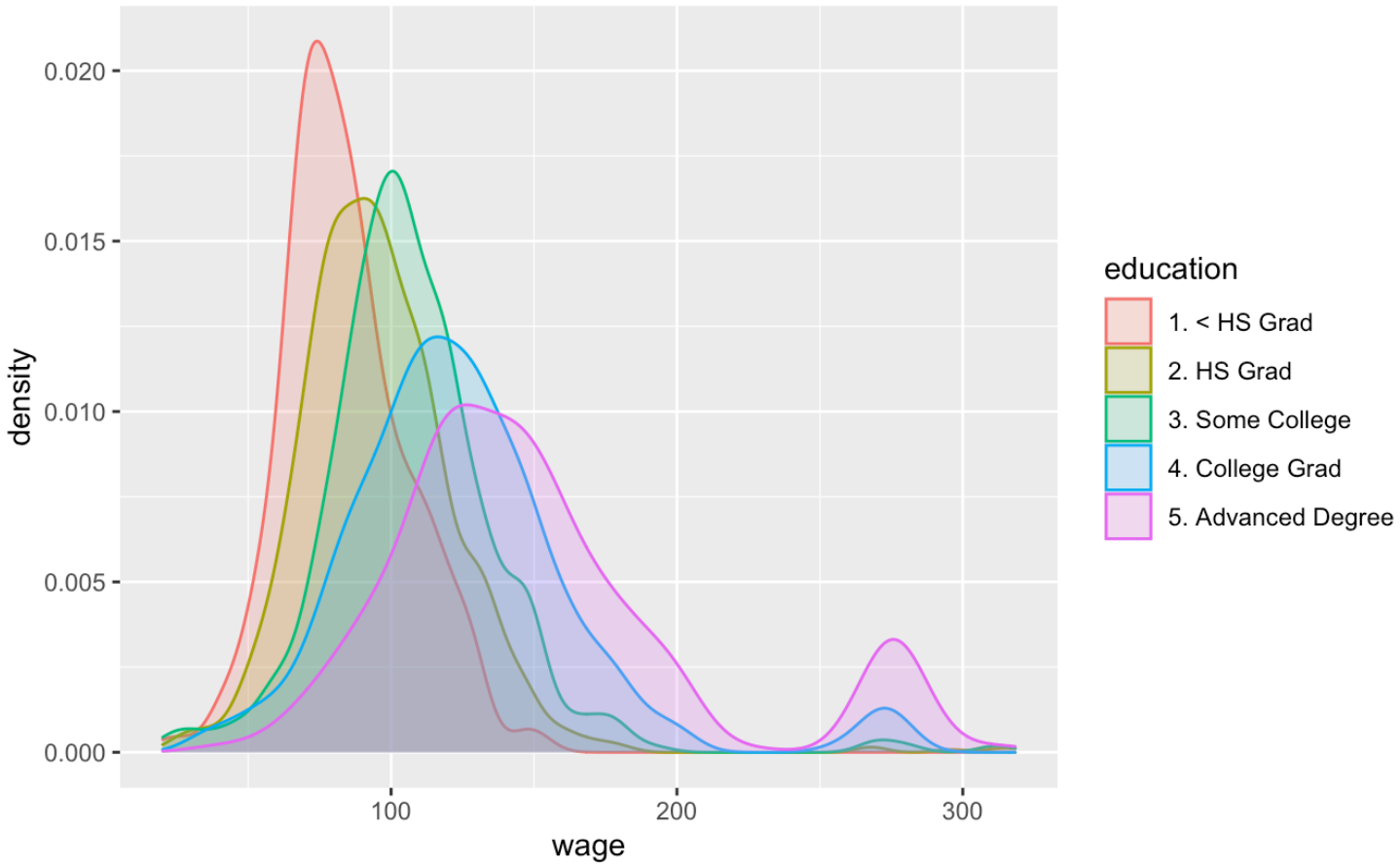
Hide

```
# overall distribution of wage by education using boxplot
ggplot(data=data, aes(y = wage, x = education, fill=education)) +
  geom_boxplot() +
  labs(title = 'Distribution of Education')
```



Hide

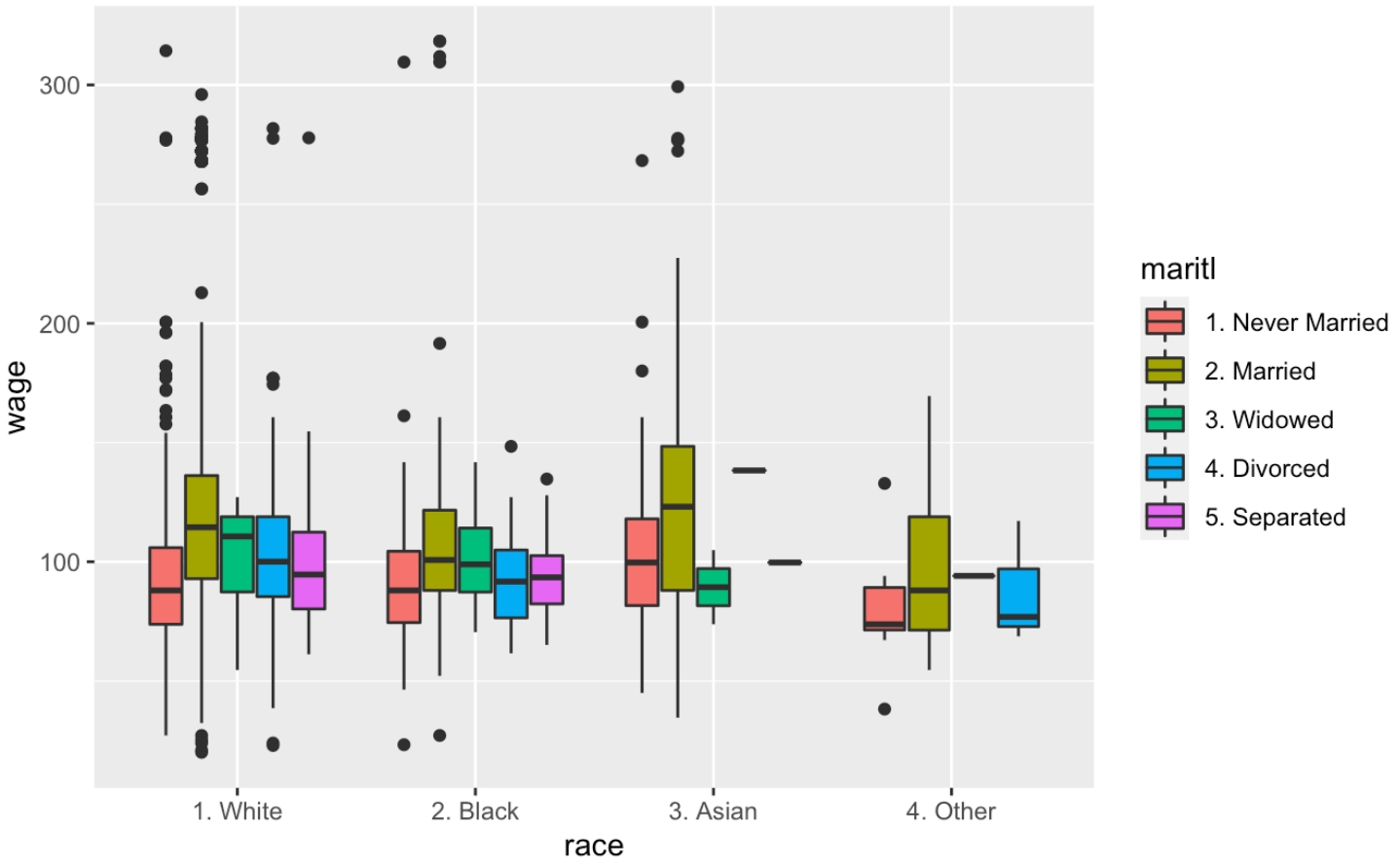
```
# overall distribution of wage by education using density plot
ggplot(data=data) +
        geom_density(aes(x = wage, color = education, fill = education) ,alpha = 0.2
)
```



Wage Distribution by Marital Status

Hide

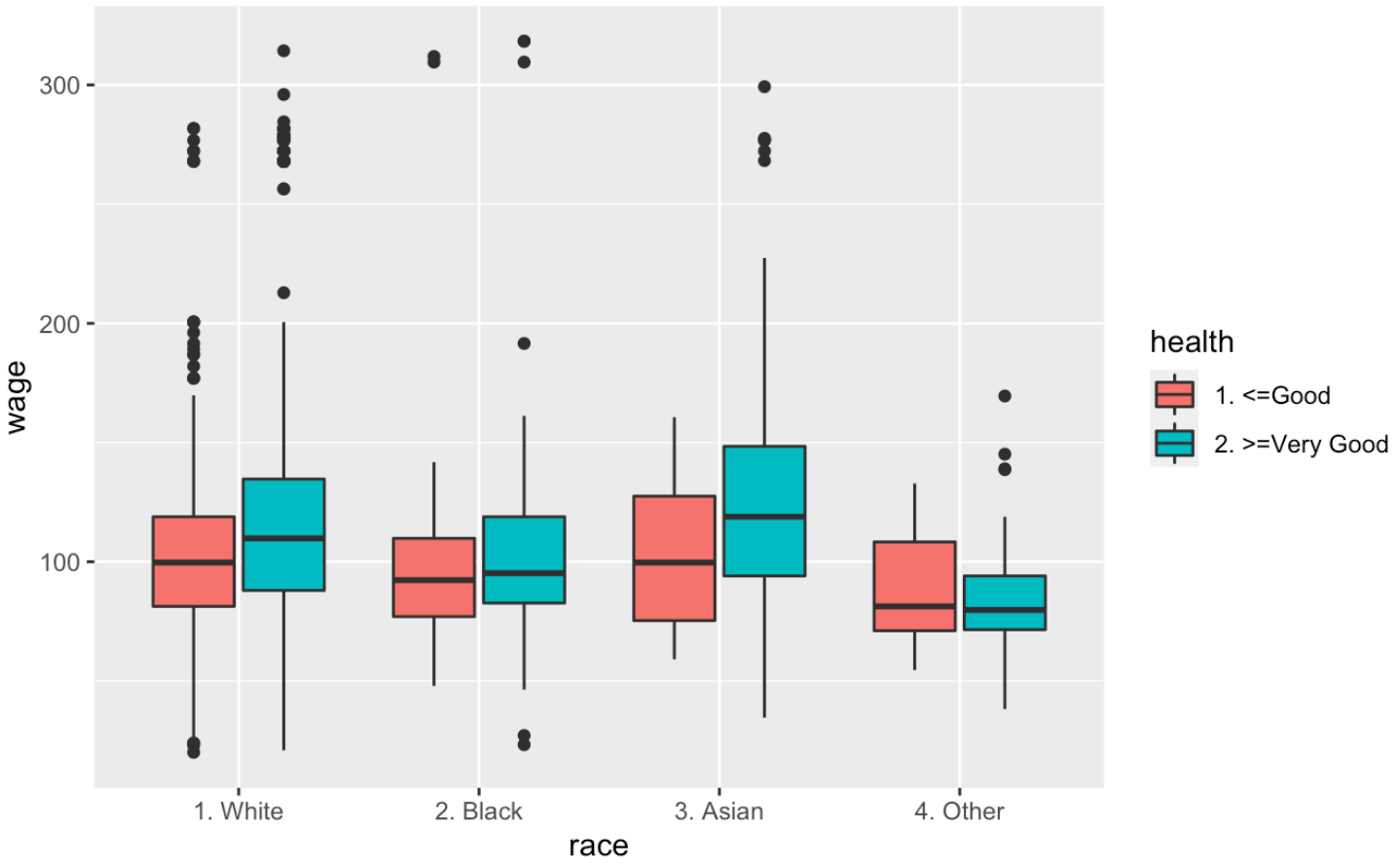
```
# overall distribution of wage by marital status using boxplot
ggplot(data=data, aes(y = wage, x = race, fill=maritl)) +
  geom_boxplot()
```



Wage Distribution by Health

Hide

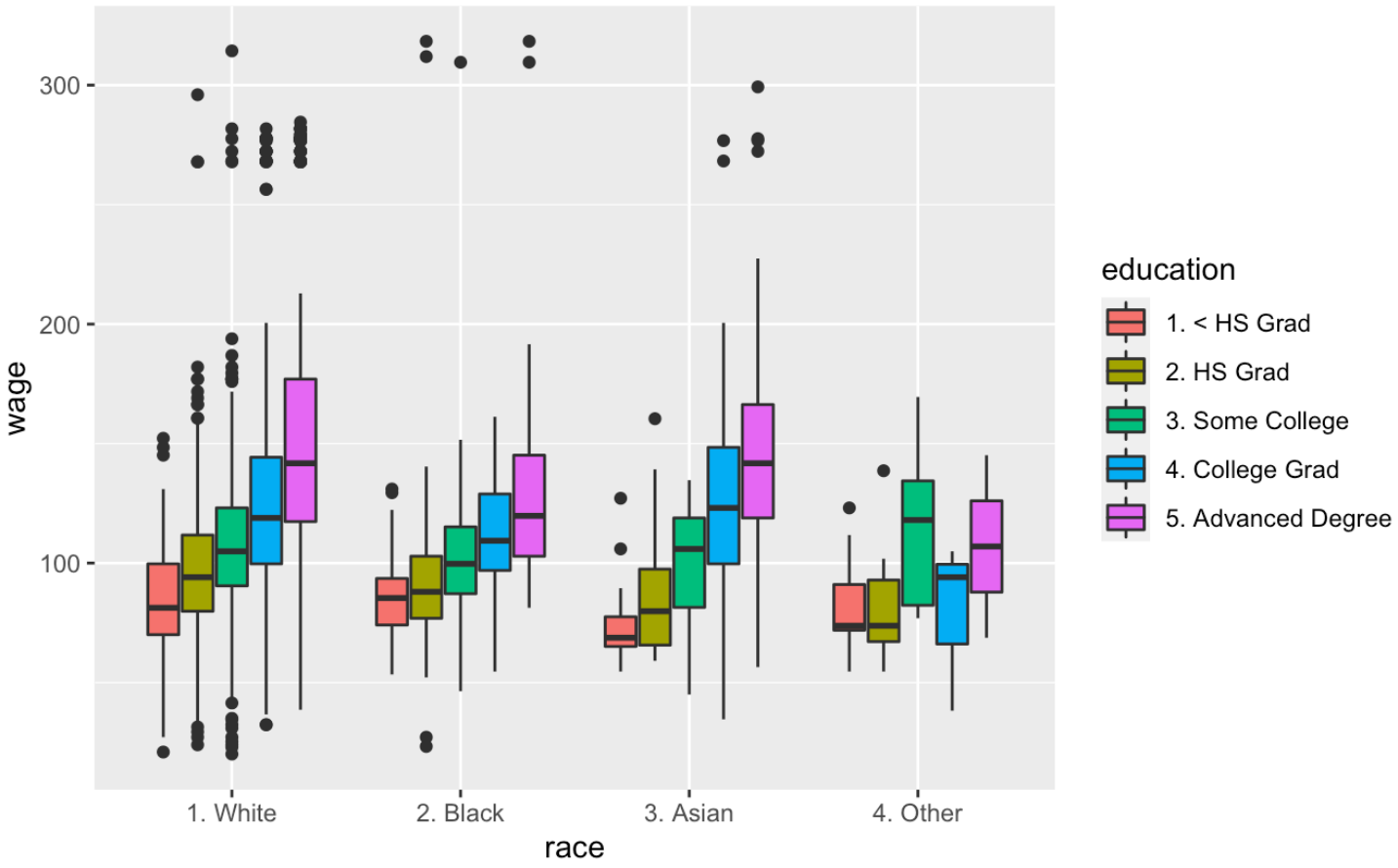
```
# overall distribution of wage by health using boxplot
ggplot(data=data, aes(y = wage, x = race, fill=health)) +
  geom_boxplot()
```



Wage Distribution by Race and Education

Hide

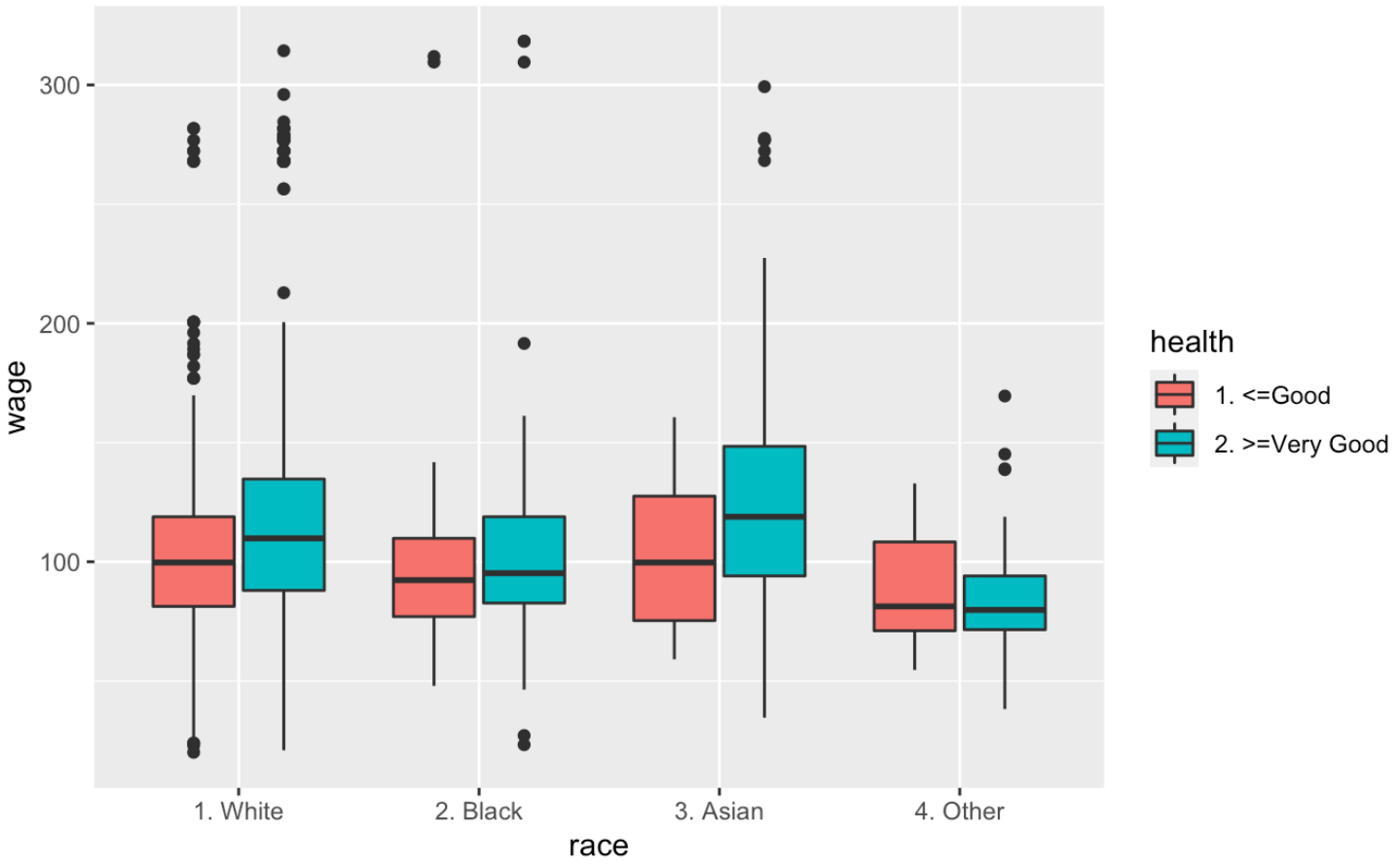
```
ggplot(data=data, aes(y = wage, x = race, fill=education)) +
  geom_boxplot()
```



Wage Distribution by Race and Health

Hide

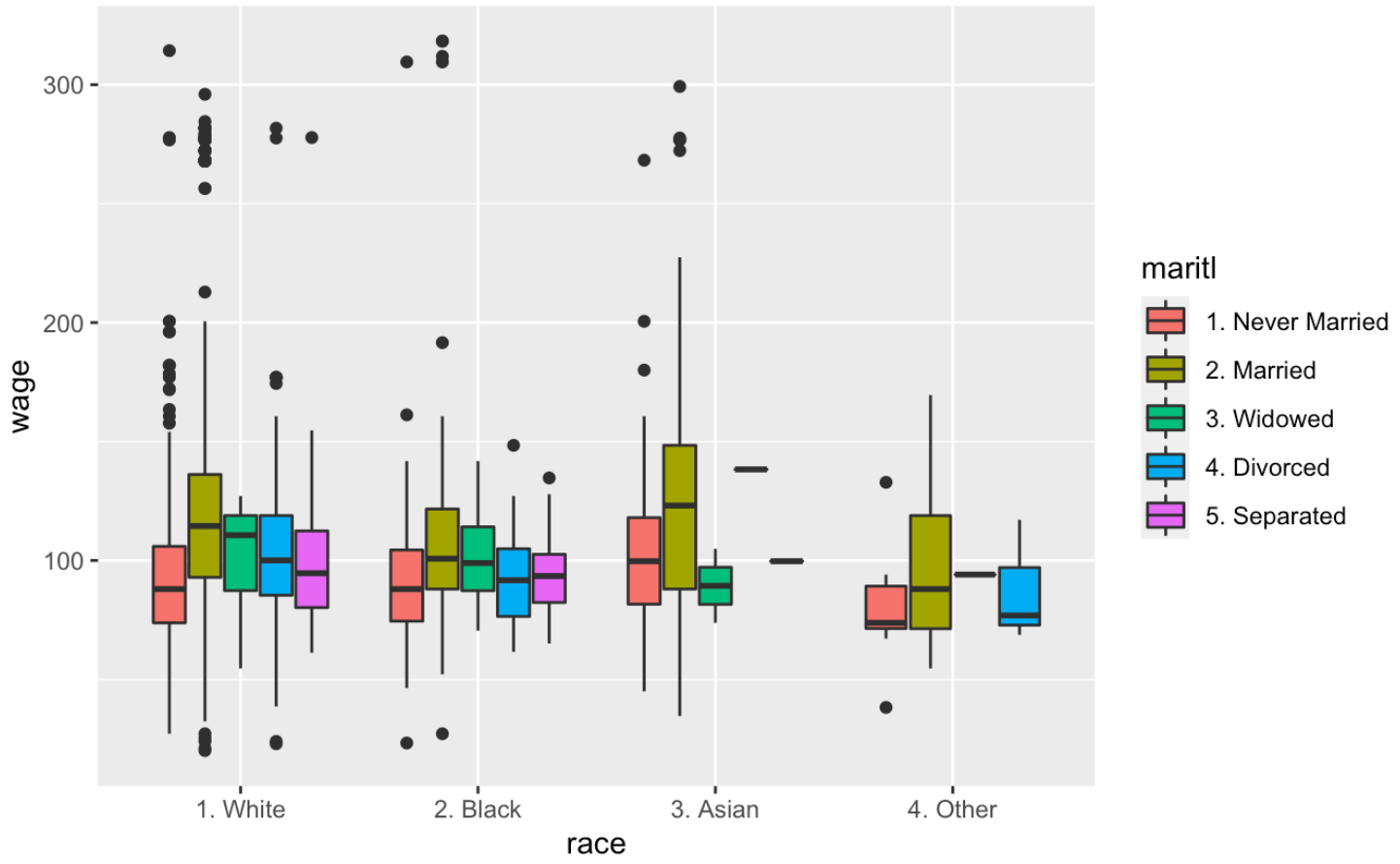
```
ggplot(data=data, aes(y = wage, x = race, fill=health)) +
  geom_boxplot()
```



Wage Distribution by Race and Marital Status

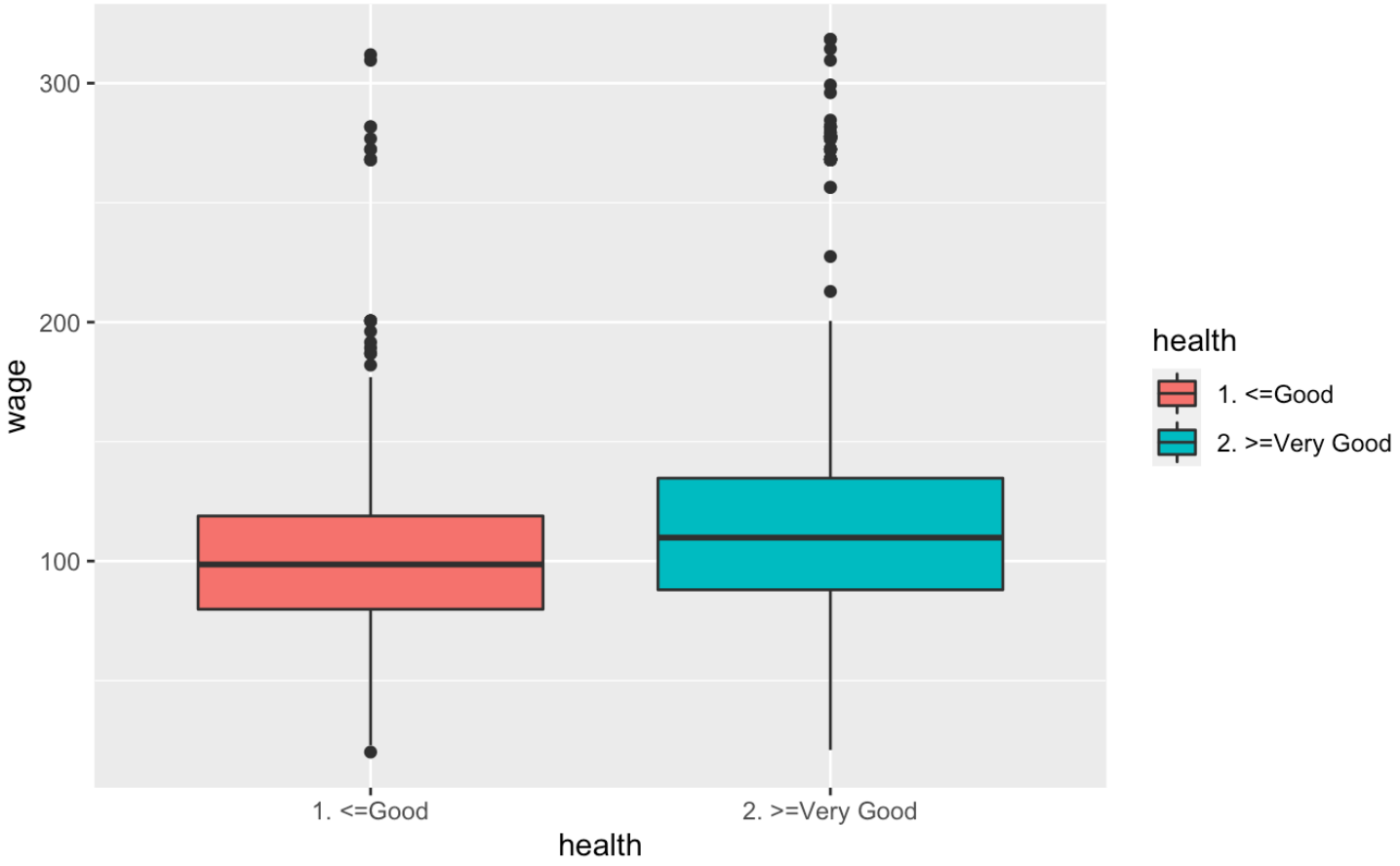
Hide

```
ggplot(data=data, aes(y = wage, x = race, fill=maritl)) +
  geom_boxplot()
```



Hide

```
ggplot(data=data, aes(y = wage, x = health, fill=health)) +
  geom_boxplot()
```



Part 2: Linear Regression

How does education, marital status, and health influence wage outcome while controlling for race?

Hide

```
# Converting health, marital status and health into factor
data$maritl.f <- as.factor(data$maritl)
data$health.f <- as.factor(data$health)
data$education.f <- as.factor(data$education)
data$race.f <- as.factor(data$race)

levels(data$maritl.f)
```

```
[1] "1. Never Married" "2. Married"       "3. Widowed"       "4. Divorced"      "5. S
eparated"
```

Hide

```
levels(data$health.f)
```

```
[1] "1. <=Good"      "2. >=Very Good"
```

Hide

```
levels(data$education.f)
```

```
[1] "1. < HS Grad"       "2. HS Grad"         "3. Some College"    "4. College Grad"
"5. Advanced Degree"
```

Hide

```
levels(data$race.f)
```

```
[1] "1. White" "2. Black" "3. Asian" "4. Other"
```

Hide

```
# Setting reference gorup
data$maritl.f <- relevel(data$maritl.f, ref = "2. Married")
data$health.f <- relevel(data$health.f, ref = "2. >=Very Good")
data$education.f <- relevel(data$education.f, ref = "2. HS Grad")
data$race.f <- relevel(data$race.f, ref = "1. White")
```

Hide

```
# Fitting the regression model
results <- lm(data$wage ~ data$maritl.f+ data$health.f + data$education.f + data$race
.f, data=data)
summary(results)
```

```
Call:
lm(formula = data$wage ~ data$maritl.f + data$health.f + data$education.f +
    data$race.f, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-115.503  -19.574   -3.456   14.287  219.541

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                       104.784      1.341  78.152  < 2e-16 ***
data$maritl.f1. Never Married     -22.168      1.599 -13.862  < 2e-16 ***
data$maritl.f3. Widowed           -14.067      8.132  -1.730   0.0838 .
data$maritl.f4. Divorced          -11.967      2.598  -4.606 4.27e-06 ***
data$maritl.f5. Separated          -5.723      4.844  -1.181   0.2375
data$health.f1. <=Good             -6.299      1.449  -4.348 1.42e-05 ***
data$education.f1. < HS Grad      -10.934      2.443  -4.475 7.92e-06 ***
data$education.f3. Some College    12.171      1.791   6.797 1.29e-11 ***
data$education.f4. College Grad    27.138      1.783  15.223  < 2e-16 ***
data$education.f5. Advanced Degree 51.819      2.096  24.721  < 2e-16 ***
data$race.f2. Black                -3.398      2.200  -1.545   0.1226
data$race.f3. Asian                -4.211      2.693  -1.564   0.1179
data$race.f4. Other                -8.100      5.866  -1.381   0.1674
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 35.23 on 2987 degrees of freedom
Multiple R-squared:  0.2902,    Adjusted R-squared:  0.2874
F-statistic: 101.8 on 12 and 2987 DF,  p-value: < 2.2e-16
```

Interpretation:

- F statistic: p-value<0.001 - The overall model is statistically significant than a model with no IV

1. Intercept: 104.784
    - Suggests that when all variables are at reference level, the average wage an a mid-Atlantic working male living in the United States for a **married, white male, who is a high school graduate, and in very good health**, makes on average 104.78 (unsure of how wage is measured)
2. Marital Status - Married
    - *Never married:* -22.168, suggesting that the average wage of males who were not married make 22.17 LESS than married men. This was found to be statistically significant from 0 at a significance level of 0.001.
    - *Widowed:* Not statistically significant from 0.

- *Divorced:* -11.97, suggesting that compared to married men, the average wage of males who were divorced make 11.97 LESS. This was found to be statistically significant from 0 at a significance level of 0.001.
- *Separated:* Not statistically significant from 0.

3. Education - HS Grad
   - *<HS Grad:* -10.93, compared to males who are HS graduates, males who completed less than HS education makes on average less than 10.93 dollars. This was found to be statistically significant from 0 at a significance level of 0.001.
   - *Some College:* 12.17, compared to males who are HS graduates, males who completed some college education make on average 12.17 more. This was found to be statistically significant from 0 at a significance level of 0.001.
   - *College Grad:* 27.138, compared to males who are HS graduates, males who completed some college education make on average 27.14 more. This was found to be statistically significant from 0 at a significance level of 0.001.
   - *Advanced Degree:* 51.819, compared to males who are HS graduates, males who completed some college education make on average 51.82 more. This was found to be statistically significant from 0 at a significance level of 0.001.
4. Race was not found to be statistic significant in predicting wage.
5. Health Status - Very good health
   - Compared to males who are in very good health, males who self-reported to be in just ‘good’ health make on average 6.30 LESS than males who were in ‘very good health’.

Due to the disproportion distribution of races in the data set, we will look at such factors with respect to each race independently in determining what decisions an individual can make in maximizing their earning capacity with respect to their educational attainment, health, and marital status, based on historical trends.

White Subset: ANOVA and Post-Hoc Tukey Test

- We first run a one-way ANOVA to see if different levels of education earn different wages
- If results are significant at an alpha significance of 0.05, there is a difference in wage when it comes to the different levels of education for white men.
- Therefore, we conduct a post-hoc Tukey HSD to determine where those differences lie for each level of education, health status, and martial status.

Hide

```
# subset for white
white = Wage[which(Wage$race=='1. White'),]
white
```

	y...	...	maritl	race	education	region	job
	<int>	<int>	<fctr>	<fctr>	<fctr>	<fctr>	<fc
231655	2006	18	1. Never Married	1. White	1. < HS Grad	2. Middle Atlantic	1. I
86582	2004	24	1. Never Married	1. White	4. College Grad	2. Middle Atlantic	2. I
161300	2003	45	2. Married	1. White	3. Some College	2. Middle Atlantic	1. I
11443	2005	50	4. Divorced	1. White	2. HS Grad	2. Middle Atlantic	2. I
376662	2008	54	2. Married	1. White	4. College Grad	2. Middle Atlantic	2. I
81404	2004	52	2. Married	1. White	2. HS Grad	2. Middle Atlantic	2. I
302778	2007	45	4. Divorced	1. White	3. Some College	2. Middle Atlantic	2. I
305706	2007	34	2. Married	1. White	2. HS Grad	2. Middle Atlantic	1. I
8690	2005	35	1. Never Married	1. White	2. HS Grad	2. Middle Atlantic	2. I
153561	2003	39	2. Married	1. White	4. College Grad	2. Middle Atlantic	1. I

1-10 of 2,480 rows | 1-8 of 11 columns Previous **1** 2 3 4 5 6 ... 100 Next

Education

Hide

```
#one-way ANOVA wage and education
one_way_anova_1 <- aov(wage~ education, data = white)
summary(one_way_anova_1)
```

```
             Df  Sum Sq Mean Sq F value Pr(>F)
education     4 1014009  253502   189.9 <2e-16 ***
Residuals  2475 3303697    1335
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# post-hoc tukey
TukeyHSD(one_way_anova_1)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = wage ~ education, data = white)

$education
                                       diff       lwr      upr     p adj
2. HS Grad-1. < HS Grad            12.25262  4.555743 19.94949 0.0001407
3. Some College-1. < HS Grad       24.56922 16.455150 32.68329 0.0000000
4. College Grad-1. < HS Grad       40.88736 32.861782 48.91294 0.0000000
5. Advanced Degree-1. < HS Grad    68.01688 59.271440 76.76232 0.0000000
3. Some College-2. HS Grad         12.31660  6.767037 17.86617 0.0000000
4. College Grad-2. HS Grad         28.63474 23.215385 34.05410 0.0000000
5. Advanced Degree-2. HS Grad      55.76426 49.326696 62.20183 0.0000000
4. College Grad-3. Some College    16.31814 10.321000 22.31528 0.0000000
5. Advanced Degree-3. Some College 43.44766 36.516678 50.37864 0.0000000
5. Advanced Degree-4. College Grad 27.12952 20.302348 33.95669 0.0000000
```

Marital Status

Hide

```
#one-way ANOVA wage and marital status
one_way_anova_2 <- aov(wage~ maritl, data = white)
summary(one_way_anova_2)
```

```
              Df  Sum Sq Mean Sq F value Pr(>F)
maritl         4  319717   79929   49.48 <2e-16 ***
Residuals   2475 3997989    1615
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# post-hoc tukey
TukeyHSD(one_way_anova_2)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = wage ~ maritl, data = white)

$maritl
                                     diff        lwr         upr     p adj
2. Married-1. Never Married    27.723710  22.171339 33.27608101 0.0000000
3. Widowed-1. Never Married     8.758257 -23.289012 40.80552628 0.9455623
4. Divorced-1. Never Married   12.424781   2.856331 21.99323112 0.0036692
5. Separated-1. Never Married  10.681554  -6.940302 28.30340944 0.4625108
3. Widowed-2. Married         -18.965453 -50.745868 12.81496255 0.4788216
4. Divorced-2. Married        -15.298929 -23.931575 -6.66628250 0.0000138
5. Separated-2. Married       -17.042156 -34.173915  0.08960212 0.0519882
4. Divorced-3. Widowed          3.666524 -29.055353 36.38840045 0.9980998
5. Separated-3. Widowed         1.923296 -33.989038 37.83563096 0.9998975
5. Separated-4. Divorced       -1.743227 -20.564055 17.07760034 0.9991006
```

Health Status

Hide

```
#one-way ANOVA wage and health status
one_way_anova_3 <- aov(wage~ health, data = white)
summary(one_way_anova_3)
```

```
              Df  Sum Sq Mean Sq F value   Pr(>F)
health         1  102351  102351   60.17 1.26e-14 ***
Residuals   2478 4215354    1701
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# post-hoc tukey
TukeyHSD(one_way_anova_3)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = wage ~ health, data = white)

$health
                             diff      lwr      upr p adj
2. >=Very Good-1. <=Good 14.29785 10.68333 17.91237     0
```

There are only two categories here so the Tukey test just tells us that white men with >=Very Good health make ~14 more than those with <=Good health.

Black Subset: ANOVA and Post-Hoc Tukey Test

Hide

```
# subset for black
black = Wage[which(Wage$race=='2. Black'),]
black
```

	y...	...	maritl	race	education	region	job
	<int>	<int>	<fctr>	<fctr>	<fctr>	<fctr>	<fc
228963	2006	41	1. Never Married	2. Black	3. Some College	2. Middle Atlantic	2. I
157244	2003	18	1. Never Married	2. Black	2. HS Grad	2. Middle Atlantic	1. I
86929	2004	39	2. Married	2. Black	2. HS Grad	2. Middle Atlantic	2. I
84377	2004	22	1. Never Married	2. Black	2. HS Grad	2. Middle Atlantic	2. I
376442	2008	21	1. Never Married	2. Black	2. HS Grad	2. Middle Atlantic	1. I
451860	2009	40	1. Never Married	2. Black	1. < HS Grad	2. Middle Atlantic	2. I
156065	2003	40	2. Married	2. Black	3. Some College	2. Middle Atlantic	2. I
449456	2009	62	2. Married	2. Black	4. College Grad	2. Middle Atlantic	2. I
156310	2003	63	4. Divorced	2. Black	3. Some College	2. Middle Atlantic	1. I
228517	2006	39	2. Married	2. Black	2. HS Grad	2. Middle Atlantic	2. I

1-10 of 293 rows | 1-8 of 11 columns Previous **1** 2 3 4 5 6 ... 30 Next

Education

Hide

```
#one-way ANOVA wage and education
one_way_anova_1 <- aov(wage~ education, data = black)
summary(one_way_anova_1)
```

```
             Df Sum Sq Mean Sq F value   Pr(>F)
education     4  50385   12596   10.28 8.55e-08 ***
Residuals   288 352881    1225
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# post-hoc tukey
TukeyHSD(one_way_anova_1)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = wage ~ education, data = black)

$education
                                       diff         lwr      upr     p adj
2. HS Grad-1. < HS Grad             6.811179 -12.8304024 26.45276 0.8761277
3. Some College-1. < HS Grad       14.565240  -5.3901530 34.52063 0.2668327
4. College Grad-1. < HS Grad       24.936656   1.9433820 47.92993 0.0260165
5. Advanced Degree-1. < HS Grad    51.009859  25.1797847 76.83993 0.0000012
3. Some College-2. HS Grad          7.754061  -5.9682430 21.47637 0.5300612
4. College Grad-2. HS Grad         18.125478   0.2712179 35.97974 0.0446533
5. Advanced Degree-2. HS Grad      44.198680  22.8146521 65.58271 0.0000003
4. College Grad-3. Some College    10.371416  -7.8275007 28.57033 0.5214953
5. Advanced Degree-3. Some College 36.444619  14.7719945 58.11724 0.0000576
5. Advanced Degree-4. College Grad 26.073202   1.5747208 50.57168 0.0305840
```

Marital Status

Hide

```
# post-hoc tukey
TukeyHSD(one_way_anova_2)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = wage ~ maritl, data = black)

$maritl
                                     diff        lwr       upr     p adj
2. Married-1. Never Married    15.896699   2.873381 28.920017 0.0080667
3. Widowed-1. Never Married     9.998040 -41.323437 61.319517 0.9836771
4. Divorced-1. Never Married    1.400021 -22.389907 25.189950 0.9998465
5. Separated-1. Never Married   3.898786 -26.905618 34.703190 0.9968593
3. Widowed-2. Married          -5.898659 -56.797194 44.999877 0.9977720
4. Divorced-2. Married        -14.496677 -37.359914  8.366559 0.4107703
5. Separated-2. Married       -11.997913 -42.092401 18.096575 0.8093548
4. Divorced-3. Widowed         -8.598019 -63.251608 46.055570 0.9927302
5. Separated-3. Widowed        -6.099255 -64.150657 51.952148 0.9984820
5. Separated-4. Divorced        2.498764 -33.584917 38.582446 0.9997071
```

Health Status

Among the black subgroup, good and very good health were not found to be statistically significant, therefore, no post-hoc Tukey test will be conducted.

Asian Subset: ANOVA and Post-Hoc Tukey Test

Hide

```
# subset for asian
asian = Wage[which(Wage$race=='3. Asian'),]
asian
```

	y…	…	maritl	race	education	region	job
	<int>	<int>	<fctr>	<fctr>	<fctr>	<fctr>	<fc
155159	2003	43	2. Married	3. Asian	4. College Grad	2. Middle Atlantic	2. I
377954	2008	30	1. Never Married	3. Asian	3. Some College	2. Middle Atlantic	2. I
160191	2003	37	1. Never Married	3. Asian	4. College Grad	2. Middle Atlantic	1. I
158226	2003	38	2. Married	3. Asian	4. College Grad	2. Middle Atlantic	2. I
81383	2004	28	2. Married	3. Asian	4. College Grad	2. Middle Atlantic	1. I
154482	2003	28	1. Never Married	3. Asian	4. College Grad	2. Middle Atlantic	2. I
81071	2004	55	2. Married	3. Asian	2. HS Grad	2. Middle Atlantic	1. I

11710	2005	47	2. Married	3. Asian	5. Advanced Degree	2. Middle Atlantic	2. I
450908	2009	70	2. Married	3. Asian	5. Advanced Degree	2. Middle Atlantic	2. I
450905	2009	43	2. Married	3. Asian	4. College Grad	2. Middle Atlantic	2. I

1-10 of 190 rows | 1-8 of 11 columns Previous **1** 2 3 4 5 6 … 19 Next

Education

Hide

```
#one-way ANOVA wage and education
one_way_anova_1 <- aov(wage~ education, data = asian)
summary(one_way_anova_1)
```

```
             Df Sum Sq Mean Sq F value   Pr(>F)
education     4 128613   32153   21.34 1.72e-14 ***
Residuals   185 278691    1506
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# post-hoc tukey
TukeyHSD(one_way_anova_1)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = wage ~ education, data = asian)

$education
                                        diff        lwr       upr     p adj
2. HS Grad-1. < HS Grad             9.978872 -23.651305  43.60905 0.9250062
3. Some College-1. < HS Grad       25.003845 -12.377231  62.38492 0.3523508
4. College Grad-1. < HS Grad       51.191999  20.607483  81.77652 0.0000726
5. Advanced Degree-1. < HS Grad    74.009772  43.143365 104.87618 0.0000000
3. Some College-2. HS Grad         15.024974 -16.660314  46.71026 0.6876256
4. College Grad-2. HS Grad         41.213128  17.931692  64.49456 0.0000227
5. Advanced Degree-2. HS Grad      64.030901  40.380368  87.68143 0.0000000
4. College Grad-3. Some College    26.188154  -2.243893  54.62020 0.0868673
5. Advanced Degree-3. Some College 49.005927  20.270865  77.74099 0.0000498
5. Advanced Degree-4. College Grad 22.817773   3.744976  41.89057 0.0102292
```

Martial Status

Hide

```
#one-way ANOVA wage and marital status
one_way_anova_2 <- aov(wage~ maritl, data = asian)
summary(one_way_anova_2)
```

```
             Df Sum Sq Mean Sq F value Pr(>F)
maritl        4  10493    2623   1.223  0.303
Residuals   185 396812    2145
```

Hide

```
# post-hoc tukey
TukeyHSD(one_way_anova_2)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = wage ~ maritl, data = asian)

$maritl
                                    diff          lwr       upr     p adj
2. Married-1. Never Married    16.068761   -7.134078  39.27160 0.3168071
3. Widowed-1. Never Married   -18.500365 -111.061894  74.06116 0.9817505
4. Divorced-1. Never Married   30.450137  -98.805004 159.70528 0.9666503
5. Separated-1. Never Married  -8.159526 -137.414667 121.09561 0.9997935
3. Widowed-2. Married         -34.569126 -125.394479  56.25623 0.8322953
4. Divorced-2. Married         14.381376 -113.636199 142.39895 0.9979926
5. Separated-2. Married       -24.228287 -152.245863 103.78929 0.9851131
4. Divorced-3. Widowed         48.950502 -107.311344 205.21235 0.9099926
5. Separated-3. Widowed        10.340839 -145.921007 166.60268 0.9997510
5. Separated-4. Divorced      -38.609663 -219.045300 141.82597 0.9765139
```

Health Status

Hide

```
#one-way ANOVA wage and health status
one_way_anova_3 <- aov(wage~ health, data = asian)
summary(one_way_anova_3)
```

```
             Df Sum Sq Mean Sq F value  Pr(>F)
health        1  18905   18905   9.151 0.00283 **
Residuals   188 388399    2066
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hide

```
# post-hoc tukey
TukeyHSD(one_way_anova_3)
```

```
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = wage ~ health, data = asian)

$health
                              diff      lwr     upr     p adj
2. >=Very Good-1. <=Good 22.51007 7.831042 37.1891 0.0028337
```