

m158-project1-sara-ian

R Markdown

```
NHANES %>%  
  glimpse()
```

```
## Rows: 10,000  
## Columns: 76  
## $ ID          <int> 51624, 51624, 51624, 51625, 51630, 51638, 51646, 5164~  
## $ SurveyYr    <fct> 2009_10, 2009_10, 2009_10, 2009_10, 2009_10, 2009_10, ~  
## $ Gender      <fct> male, male, male, male, female, male, male, female, f~  
## $ Age         <int> 34, 34, 34, 4, 49, 9, 8, 45, 45, 45, 66, 58, 54, 10, ~  
## $ AgeDecade   <fct> 30-39, 30-39, 30-39, 0-9, 40-49, 0-9, 0-9, 40~  
## $ AgeMonths   <int> 409, 409, 409, 49, 596, 115, 101, 541, 541, 541, 795, ~  
## $ Race1       <fct> White, White, White, Other, White, White, White, Whit~  
## $ Race3       <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~  
## $ Education   <fct> High School, High School, High School, NA, Some Colle~  
## $ MaritalStatus <fct> Married, Married, Married, NA, LivePartner, NA, NA, M~  
## $ HHIncome    <fct> 25000-34999, 25000-34999, 25000-34999, 20000-24999, 3~  
## $ HHIncomeMid <int> 30000, 30000, 30000, 22500, 40000, 87500, 60000, 8750~  
## $ Poverty     <dbl> 1.36, 1.36, 1.36, 1.07, 1.91, 1.84, 2.33, 5.00, 5.00, ~  
## $ HomeRooms   <int> 6, 6, 6, 9, 5, 6, 7, 6, 6, 5, 10, 6, 10, 10, 4, 3, ~  
## $ HomeOwn     <fct> Own, Own, Own, Own, Rent, Rent, Own, Own, Own, Own, O~  
## $ Work        <fct> NotWorking, NotWorking, NotWorking, NA, NotWorking, N~  
## $ Weight      <dbl> 87.4, 87.4, 87.4, 17.0, 86.7, 29.8, 35.2, 75.7, 75.7, ~  
## $ Length      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~  
## $ HeadCirc    <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~  
## $ Height      <dbl> 164.7, 164.7, 164.7, 105.4, 168.4, 133.1, 130.6, 166.~  
## $ BMI         <dbl> 32.22, 32.22, 32.22, 15.30, 30.57, 16.82, 20.64, 27.2~  
## $ BMICatUnder20yrs <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~  
## $ BMI_WHO     <fct> 30.0_plus, 30.0_plus, 30.0_plus, 12.0_18.5, 30.0_plus~  
## $ Pulse       <int> 70, 70, 70, NA, 86, 82, 72, 62, 62, 60, 62, 76, 8~  
## $ BPSysAve    <int> 113, 113, 113, NA, 112, 86, 107, 118, 118, 118, 111, ~  
## $ BPDiaAve    <int> 85, 85, 85, NA, 75, 47, 37, 64, 64, 64, 63, 74, 85, 6~  
## $ BPSys1      <int> 114, 114, 114, NA, 118, 84, 114, 106, 106, 106, 124, ~  
## $ BPDia1      <int> 88, 88, 88, NA, 82, 50, 46, 62, 62, 62, 64, 76, 86, 6~  
## $ BPSys2      <int> 114, 114, 114, NA, 108, 84, 108, 118, 118, 118, 108, ~  
## $ BPDia2      <int> 88, 88, 88, NA, 74, 50, 36, 68, 68, 68, 62, 72, 88, 6~  
## $ BPSys3      <int> 112, 112, 112, NA, 116, 88, 106, 118, 118, 118, 114, ~  
## $ BPDia3      <int> 82, 82, 82, NA, 76, 44, 38, 60, 60, 60, 64, 76, 82, 7~  
## $ Testosterone <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~  
## $ DirectChol  <dbl> 1.29, 1.29, 1.29, NA, 1.16, 1.34, 1.55, 2.12, 2.12, 2~  
## $ TotChol     <dbl> 3.49, 3.49, 3.49, NA, 6.70, 4.86, 4.09, 5.82, 5.82, 5~  
## $ UrineVol1   <int> 352, 352, 352, NA, 77, 123, 238, 106, 106, 106, 113, ~  
## $ UrineFlow1  <dbl> NA, NA, NA, NA, 0.094, 1.538, 1.322, 1.116, 1.116, 1.~
```

```
## $ UrineVol2      <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ UrineFlow2    <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ Diabetes      <fct> No, No, No, No, No, No, No, No, No, No, No, No, No, No, No, N~
## $ DiabetesAge   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ HealthGen     <fct> Good, Good, Good, NA, Good, NA, NA, Vgood, Vgood, Vgo~
## $ DaysPhysHlthBad <int> 0, 0, 0, NA, 0, NA, NA, 0, 0, 0, 10, 0, 4, NA, NA, 0, ~
## $ DaysMentHlthBad <int> 15, 15, 15, NA, 10, NA, NA, 3, 3, 3, 0, 0, 0, NA, NA, ~
## $ LittleInterest <fct> Most, Most, Most, NA, Several, NA, NA, None, None, No~
## $ Depressed     <fct> Several, Several, Several, NA, Several, NA, NA, None, ~
## $ nPregnancies  <int> NA, NA, NA, NA, 2, NA, NA, 1, 1, 1, NA, NA, NA, NA, N~
## $ nBabies       <int> NA, NA, NA, NA, 2, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
## $ Age1stBaby    <int> NA, NA, NA, NA, 27, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ SleepHrsNight <int> 4, 4, 4, NA, 8, NA, NA, 8, 8, 8, 7, 5, 4, NA, 5, 7, N~
## $ SleepTrouble  <fct> Yes, Yes, Yes, NA, Yes, NA, NA, No, No, No, No, No, No, Y~
## $ PhysActive    <fct> No, No, No, NA, No, NA, NA, Yes, Yes, Yes, Yes, Yes, ~
## $ PhysActiveDays <int> NA, NA, NA, NA, NA, NA, NA, 5, 5, 5, 7, 5, 1, NA, 2, ~
## $ TVHrsDay      <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ CompHrsDay    <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
## $ TVHrsDayChild <int> NA, NA, NA, 4, NA, 5, 1, NA, NA, NA, NA, NA, NA, NA, 4, N~
## $ CompHrsDayChild <int> NA, NA, NA, 1, NA, 0, 6, NA, NA, NA, NA, NA, NA, NA, 3, N~
## $ Alcohol12PlusYr <fct> Yes, Yes, Yes, NA, Yes, NA, NA, Yes, Yes, Yes, Yes, Y~
## $ AlcoholDay    <int> NA, NA, NA, NA, 2, NA, NA, 3, 3, 3, 1, 2, 6, NA, NA, ~
## $ AlcoholYear   <int> 0, 0, 0, NA, 20, NA, NA, 52, 52, 52, 100, 104, 364, N~
## $ SmokeNow      <fct> No, No, No, NA, Yes, NA, NA, NA, NA, NA, NA, No, NA, NA, ~
## $ Smoke100      <fct> Yes, Yes, Yes, NA, Yes, NA, NA, No, No, No, Yes, No, ~
## $ Smoke100n     <fct> Smoker, Smoker, Smoker, NA, Smoker, NA, NA, Non-Smoke~
## $ SmokeAge      <int> 18, 18, 18, NA, 38, NA, NA, NA, NA, NA, 13, NA, NA, N~
## $ Marijuana     <fct> Yes, Yes, Yes, NA, Yes, NA, NA, Yes, Yes, Yes, NA, Ye~
## $ AgeFirstMarij <int> 17, 17, 17, NA, 18, NA, NA, 13, 13, 13, NA, 19, 15, N~
## $ RegularMarij  <fct> No, No, No, NA, No, NA, NA, No, No, No, NA, Yes, Yes, ~
## $ AgeRegMarij   <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 20, 15, N~
## $ HardDrugs     <fct> Yes, Yes, Yes, NA, Yes, NA, NA, No, No, No, No, Yes, ~
## $ SexEver       <fct> Yes, Yes, Yes, NA, Yes, NA, NA, Yes, Yes, Yes, Yes, Y~
## $ SexAge        <int> 16, 16, 16, NA, 12, NA, NA, 13, 13, 13, 17, 22, 12, N~
## $ SexNumPartnLife <int> 8, 8, 8, NA, 10, NA, NA, 20, 20, 20, 15, 7, 100, NA, ~
## $ SexNumPartYear <int> 1, 1, 1, NA, 1, NA, NA, 0, 0, 0, NA, 1, 1, NA, NA, 1, ~
## $ SameSex       <fct> No, No, No, NA, Yes, NA, NA, Yes, Yes, Yes, No, No, N~
## $ SexOrientation <fct> Heterosexual, Heterosexual, Heterosexual, NA, Heteros~
## $ PregnantNow   <fct> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, N~
```

Varibales we're using: Age (quantitative) Income (numerical discrete?) Testosterone (quantitative) Physactivedays (numerical discrete) Pulse (quantitative)

Gender (categorical) SleepTrouble (categorical) Depressed (categorical) Education (categorical) Work (categorical)

```
OurData <- NHANES %>% select(Age, HHIncome, Testosterone, PhysActiveDays,
                             Pulse, Gender, SleepTrouble, Depressed,
                             Education, Work)
names(OurData)
```

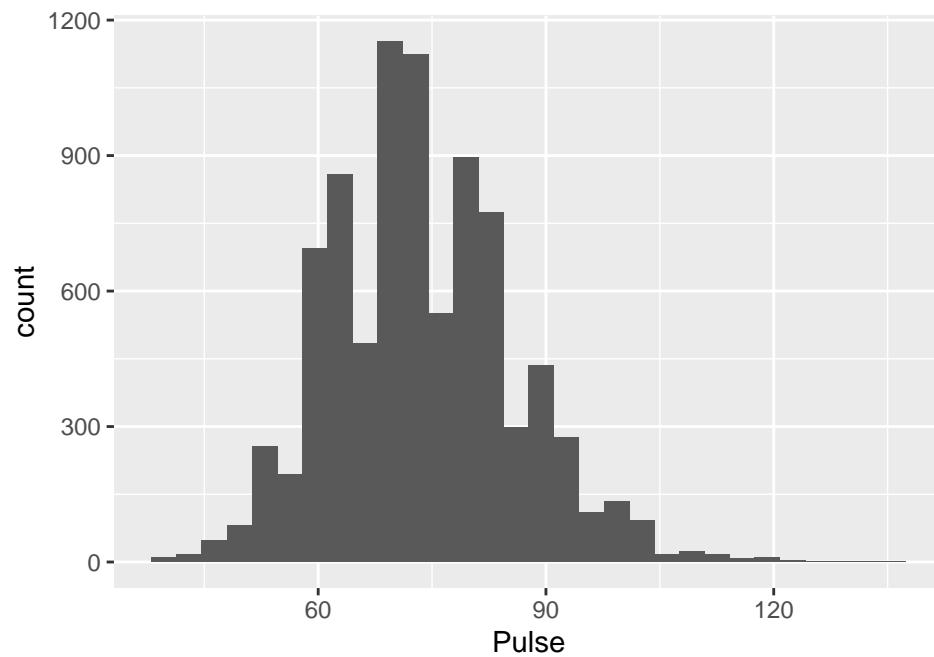
```
## [1] "Age"           "HHIncome"      "Testosterone"  "PhysActiveDays"
## [5] "Pulse"         "Gender"        "SleepTrouble"  "Depressed"
## [9] "Education"     "Work"
```

```
dim(OurData)
```

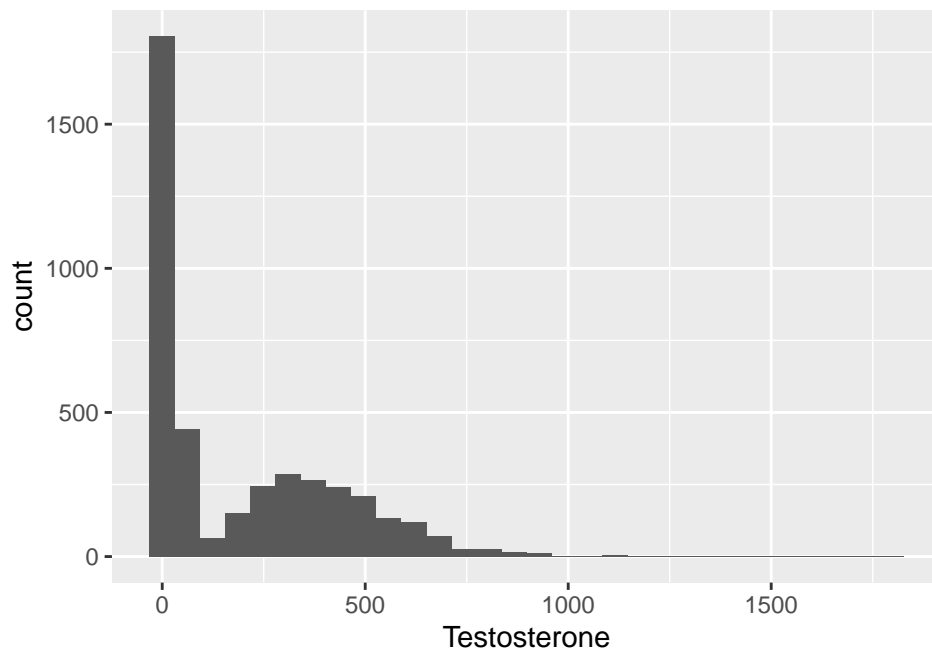
```
## [1] 10000    10
```

```
skim(OurData)
```

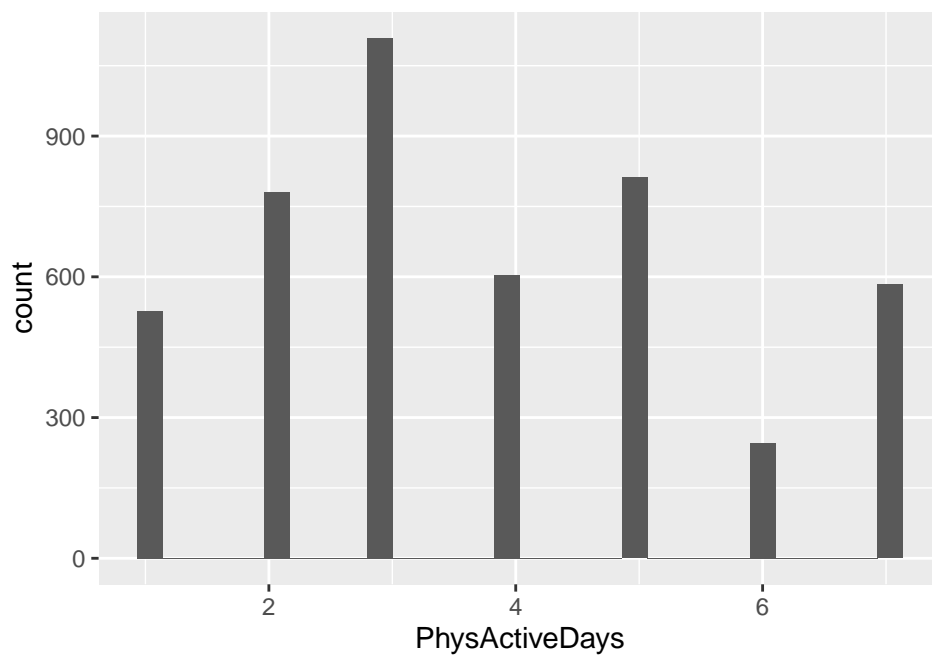
```
OurData %>%  
  ggplot(aes(x = Pulse)) + # which column are the residuals?  
  geom_histogram() # nothing to do here, but remember this function for later!
```



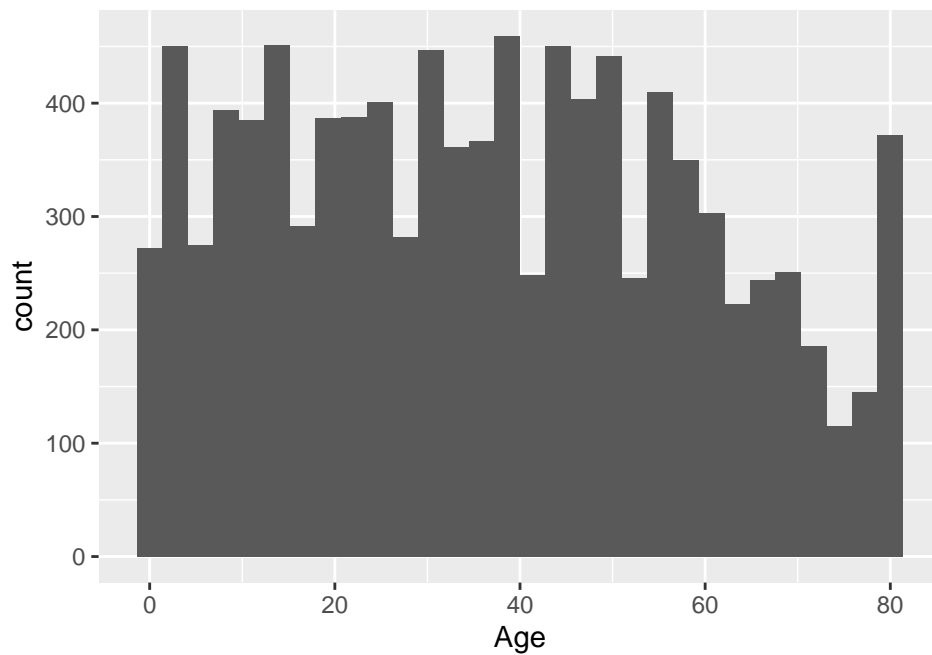
```
OurData %>%  
  ggplot(aes(x = Testosterone)) + # which column are the residuals?  
  geom_histogram() # nothing to do here, but remember this function for later!
```



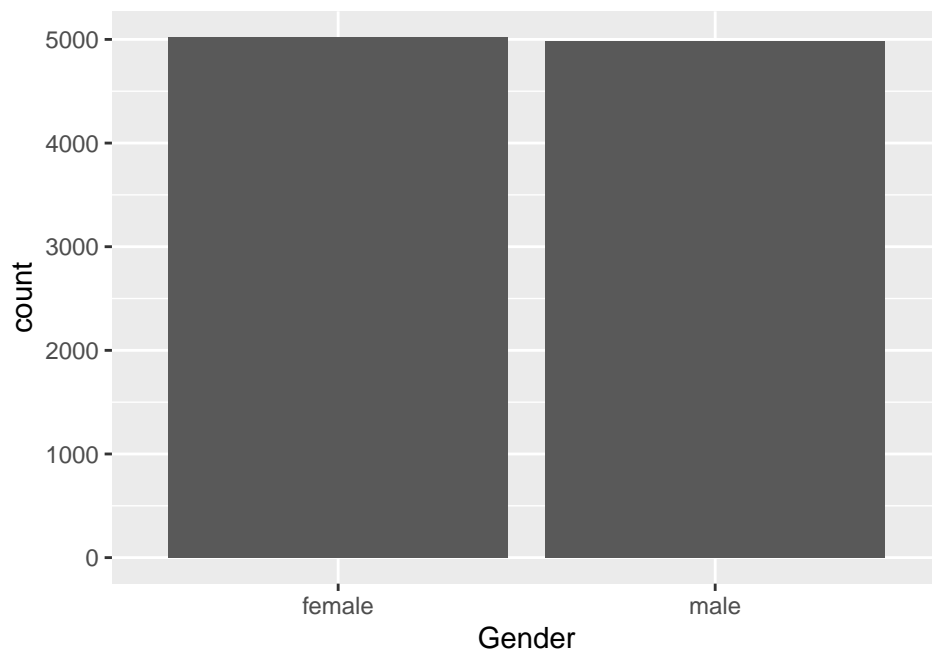
```
OurData %>%
  ggplot(aes(x = PhysActiveDays)) + # which column are the residuals?
  geom_histogram() # nothing to do here, but remember this function for later!
```



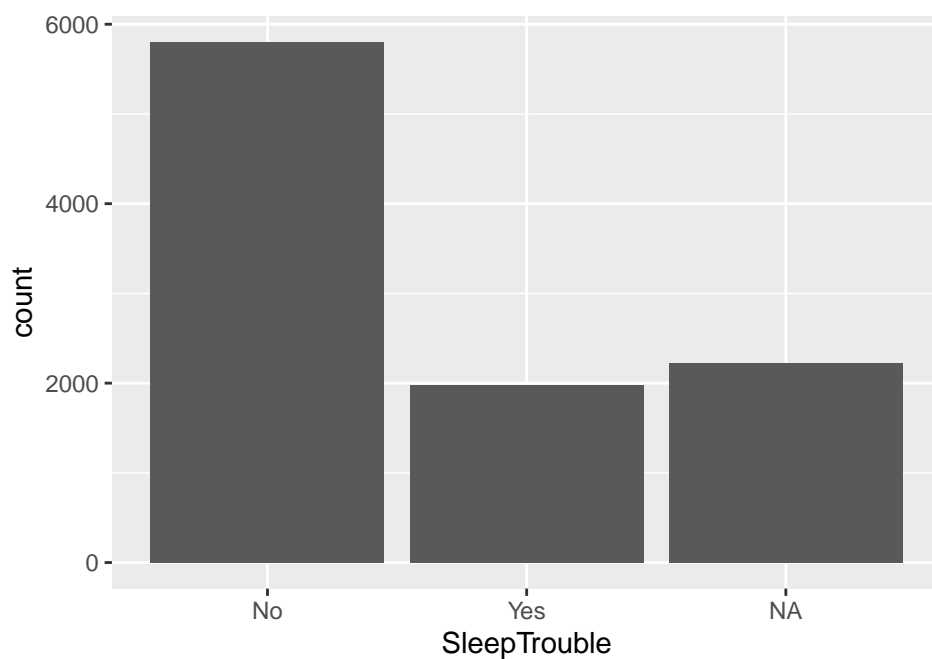
```
OurData %>%
  ggplot(aes(x = Age)) + # which column are the residuals?
  geom_histogram() # nothing to do here, but remember this function for later!
```



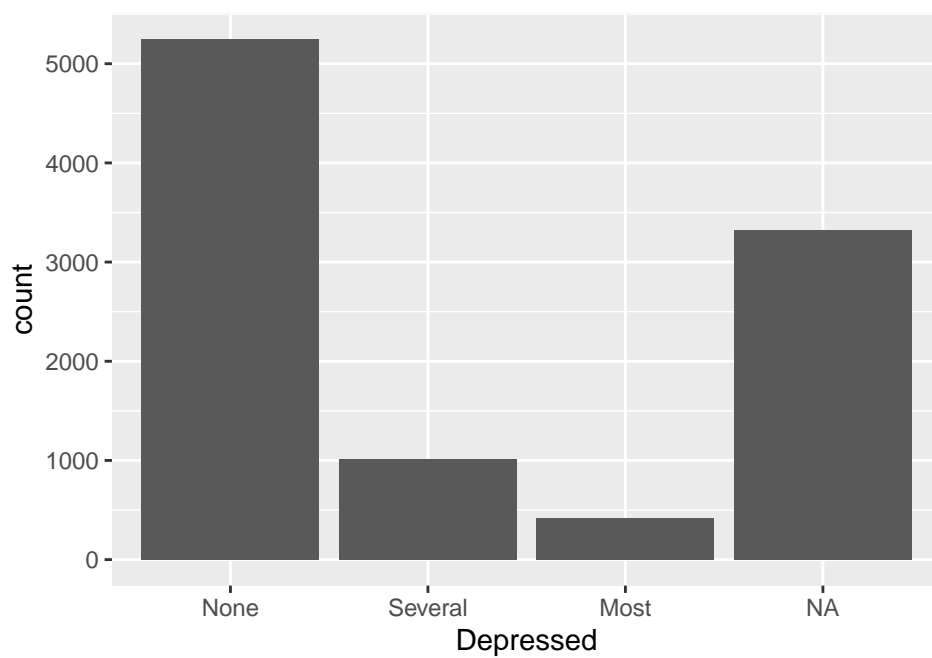
```
OurData %>%
  ggplot(aes(x = Gender)) + # which column are the residuals?
  geom_bar() # nothing to do here, but remember this function for later!
```



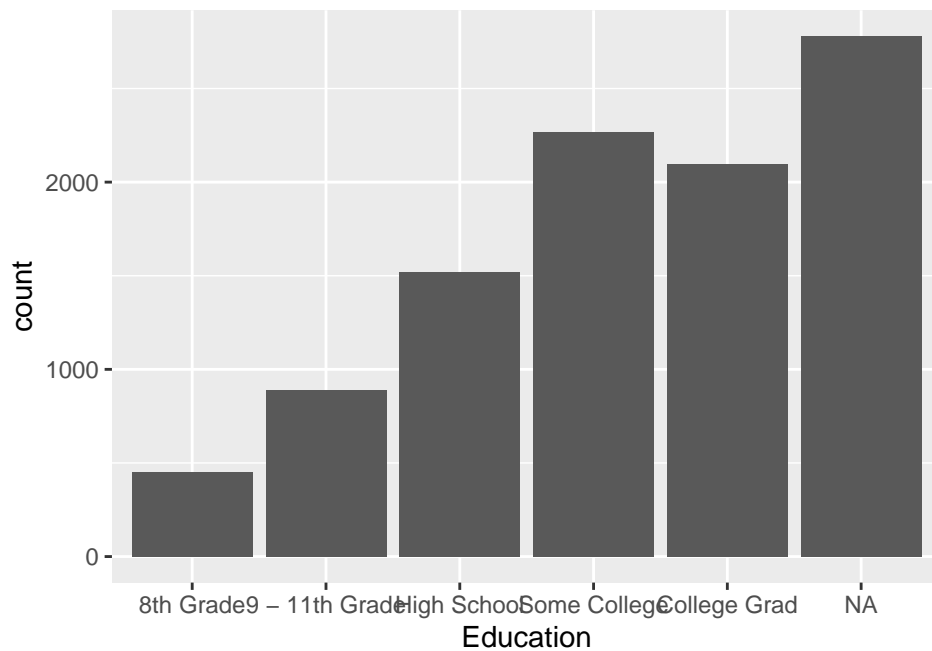
```
OurData %>%
  ggplot(aes(x = SleepTrouble)) + # which column are the residuals?
  geom_bar() # nothing to do here, but remember this function for later!
```



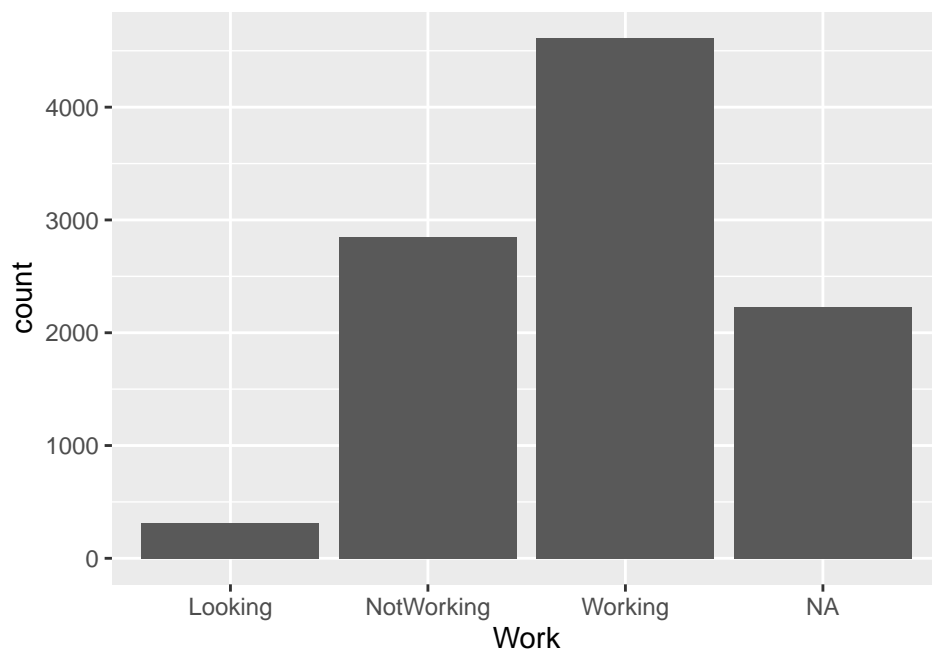
```
OurData %>%
  ggplot(aes(x = Depressed)) + # which column are the residuals?
  geom_bar() # nothing to do here, but remember this function for later!
```



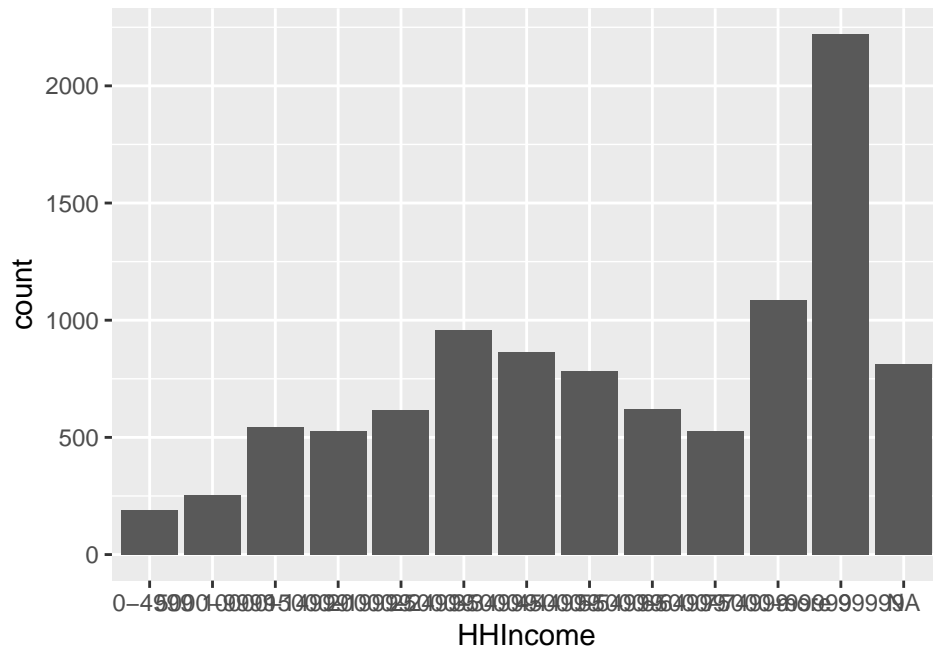
```
OurData %>%
  ggplot(aes(x = Education)) + # which column are the residuals?
  geom_bar() # nothing to do here, but remember this function for later!
```



```
OurData %>%
  ggplot(aes(x = Work)) + # which column are the residuals?
  geom_bar() # nothing to do here, but remember this function for later!
```



```
OurData %>%
  ggplot(aes(x = HHIncome)) + # which column are the residuals?
  geom_bar() # nothing to do here, but remember this function for later!
```



Change We are using data from the National Health and Nutrition Exam Survey (NHANES) database that was collected between 2009-2012 with adjusted weighing.

There are 75 total variables, however only 10 are relevant to our project.

The target population of NHANES is “the non-institutionalized civilian resident population of the United States”. Therefore, the observational unit is a civilian resident of the United States (of any age).