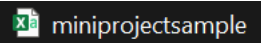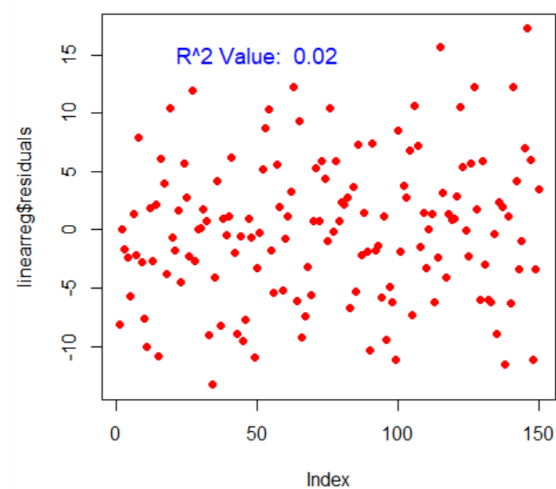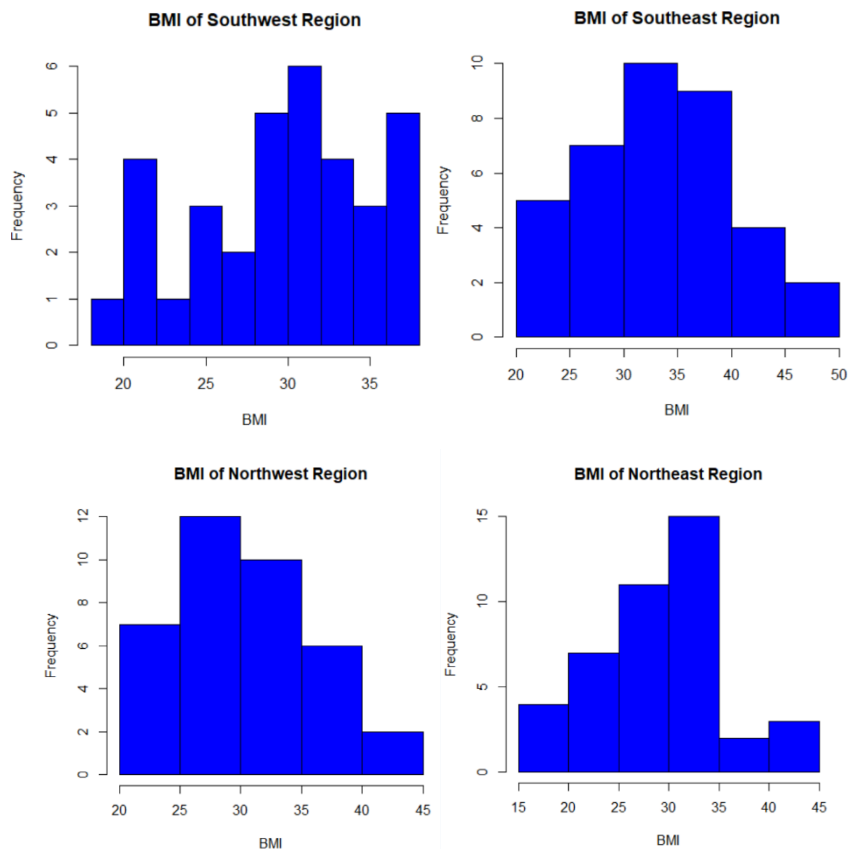1. Randomly selected 150 rows and creating/exporting new csv

   XII miniprojectsample

2. Finding R^2 (linear relationship) between bmi and charges and plotting the linear regression


3. Create data frame containing only the BMI column in each region stated in the column, "Region," ie. southwest, southeast, northwest, northeast, and create a histogram

4. Calculate average charges and max/min charges for each region from the reduced data and visualize the regional disparities (region vs charges) in a bar plot

```
Max/Min charge ($) for each region:
      region charges.max charges.min
1 northeast    39047.285     1984.453
2 northwest    42760.502     2117.339
3 southeast    46889.261     1621.883
4 southwest    42856.838     1731.677
Average charges ($) for each region:
      region   charges
1 northeast 10328.75
2 northwest 13589.00
3 southeast 15883.81
4 southwest 11847.12
```


Regional Disparities
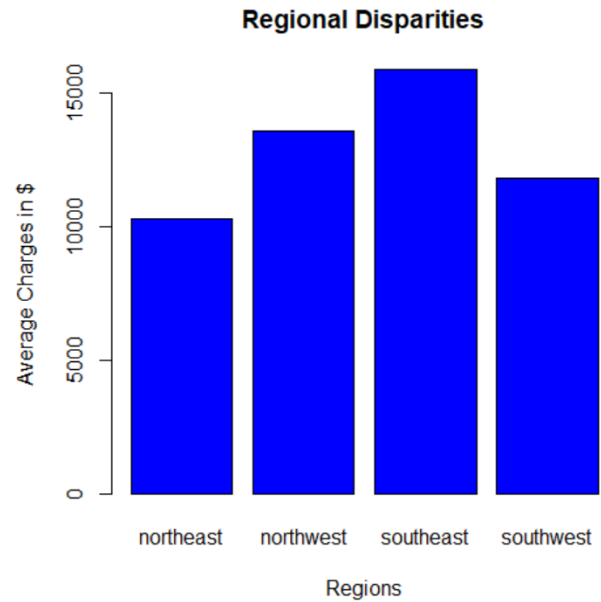
5. Create bins that categorize the BMI and calculate average charges for each category with original dataset

```
Average Charge ($) for Each BMI Category:
   bmicategories    charges
1    Underweight  8657.621
2         Normal 10404.900
3     Overweight 11006.810
4        Obese I 14217.617
5       Obese II 17245.410
6      Obese III 16667.608
```

6. Find standard deviation of the charges for each category and visualize the charges with the BMI bins

```
Standard Deviation for Each BMI Category:
   BMICategories SDCharges
1    Underweight  7591.730
2         Normal  7508.166
3     Overweight  8004.176
4        Obese I 13391.566
5       Obese II 15344.209
6      Obese III 16496.997
```

BMI vs Avg Charges $
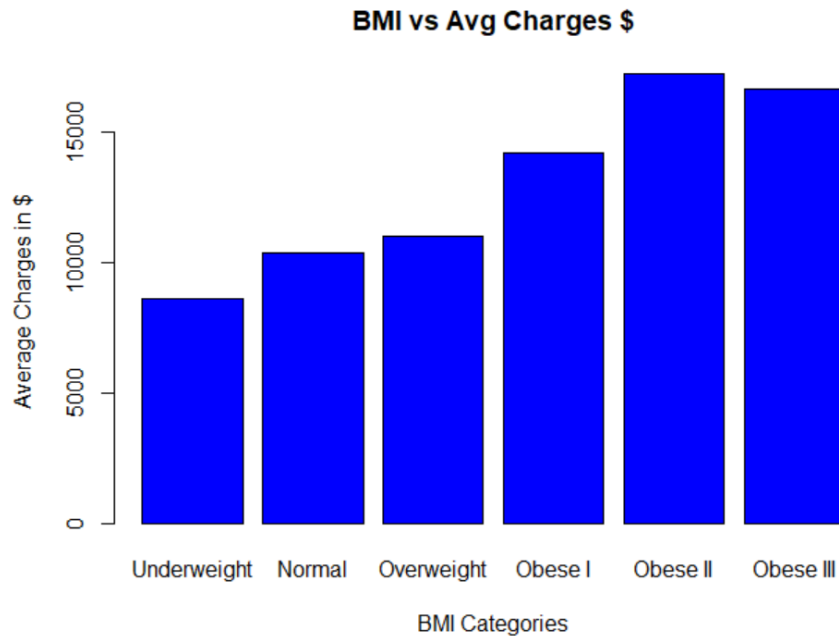
7. Calculate average charges between smokers and nonsmokers in the original dataset vs reduced dataset

```
Average Charges ($) of (non)smokers from original dataset:
  smoker    charges
1     no   8434.268
2    yes  32050.232
Average Charges ($) of (non)smokers from sample dataset:
  smoker    charges
1     no   8239.633
2    yes  27987.016
```

8. Find outliers in the dataset for bmi and charges group and plot them as a boxplot

```
Outliers in BMI:
[1] 46.09 47.52
Outliers in Charges:
 [1] 38746.36 39047.29 45008.96 39722.75 46889.26 42560.43 42760.50 40720.55
 [9] 37607.53 46599.11 42856.84 37079.37 44202.65
```

BMI Outliers

Charges Outliers
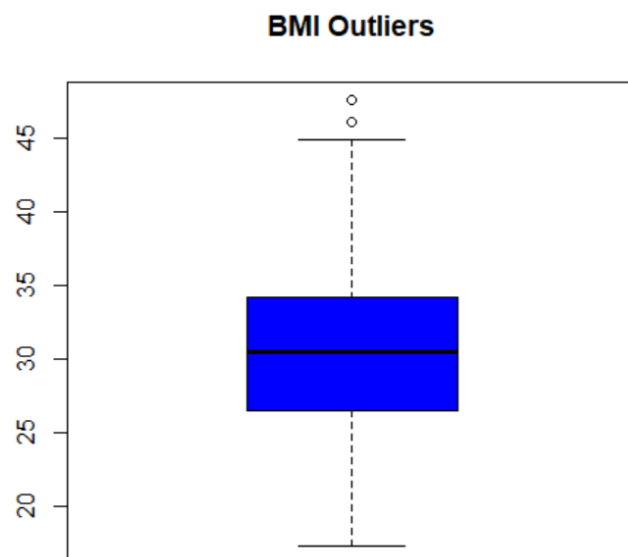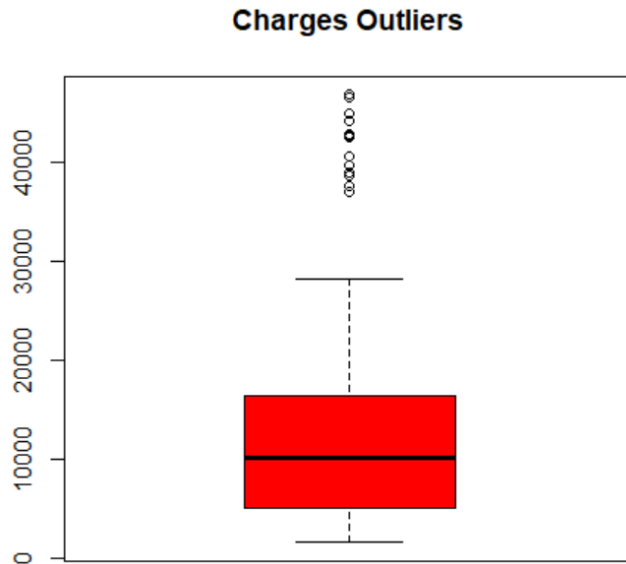
9. Find average charges of all smokers (yes) who are male from the original vs reduced data set

```
Average charges ($) of all smokers who are male from original dataset
[1] 33042.01
Average charges ($) of all smokers who are male from sample dataset
[1] 31030.29
```

10. Find average charges of individuals with <23 BMI in each of the four regions from the reduced dataset and create a bar plot with a legend to visualize this.

```
Average charges ($) of BMI <23 in southeast region from reduced dataset:
[1] 19981.78
Average charges ($) of BMI <23 in southwest region from reduced dataset:
[1] 10450.85
Average charges ($) of BMI <23 in northeast region from reduced dataset:
[1] 6218.817
Average charges ($) of BMI <23 in northwest region from reduced dataset:
[1] 9021.738
```

Average Charges from Individuals w/ < 23 BMI in Regions: SE, SW, NE, NW