Kunal Kolhe - 112748535

Report for Project 1 of Visualization course

The dataset

I used the Statlog (Australian Credit Approval) Data Set from the UCI Machine Learning Repository.

This data has a good mix of categorical and continuous variables. It has 690 entries, but I trimmed the dataset down to 259 entries. The trimming was done randomly.

Each data entry is comprised of 15 attributes.

There are 6 numerical and 8 categorical attributes. The labels have been changed for the convenience of the statistical algorithms. For example, attribute 4 originally had 3 labels p,g,gg and these have been changed to labels 1,2,3.

A1: 0,1 CATEGORICAL (formerly: a,b)
A2: continuous.
A3: continuous.
A4: 1,2,3 CATEGORICAL (formerly: p,g,gg)
A5: 1, 2,3,4,5, 6,7,8,9,10,11,12,13,14 CATEGORICAL (formerly: ff,d,i,k,j,aa,m,c,w, e, q, r,cc, x)
A6: 1, 2,3, 4,5,6,7,8,9 CATEGORICAL (formerly: ff,dd,j,bb,v,n,o,h,z)
A7: continuous.
A8: 1, 0 CATEGORICAL (formerly: t, f)
A9: 1, 0 CATEGORICAL (formerly: t, f)
A10: continuous.
A11: 1, 0 CATEGORICAL (formerly t, f)
A12: 1, 2, 3 CATEGORICAL (formerly: s, g, p)
A13: continuous.
A14: continuous.
A15: 1,2 class attribute (formerly: +,-)

A15 is actually the label but I used it as a categorical variable for the visualization.

The range of some of the attributes is quite small so I multiplied some of the attributes by multiples of 10 to get them in a good range for visualization.

I made bar charts for Categorical variables and histograms for Continuous variables. The continuous variables have sliders for adjusting the number of bins to get better granularity.
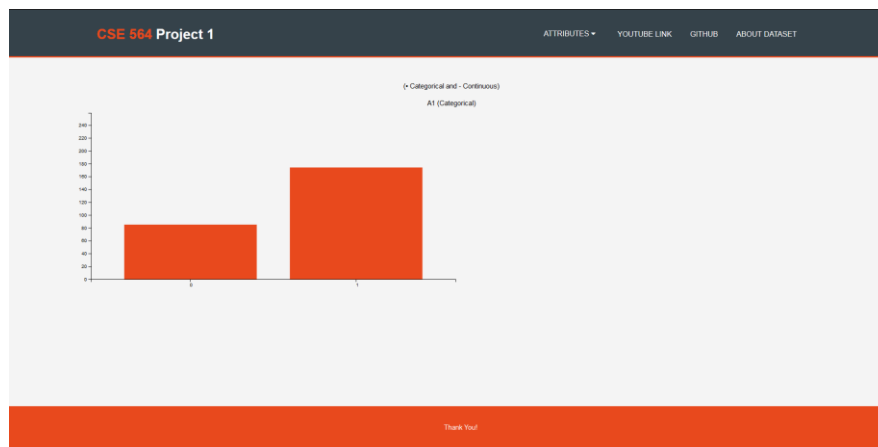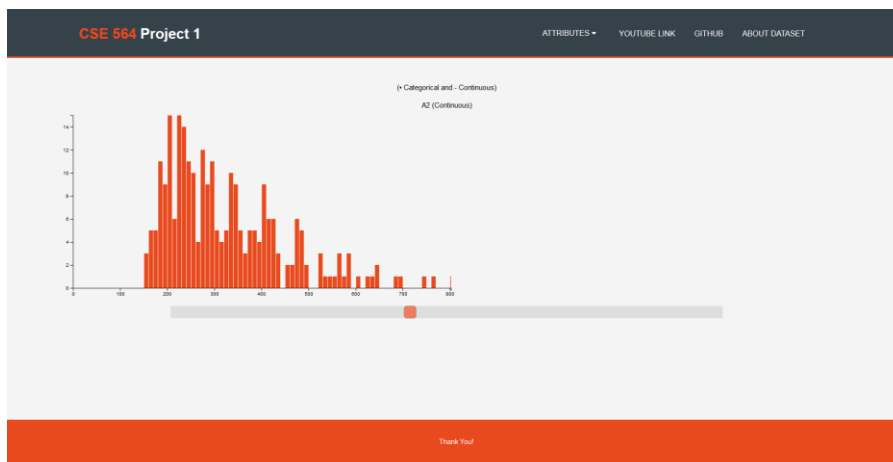
A categorical variable A1.
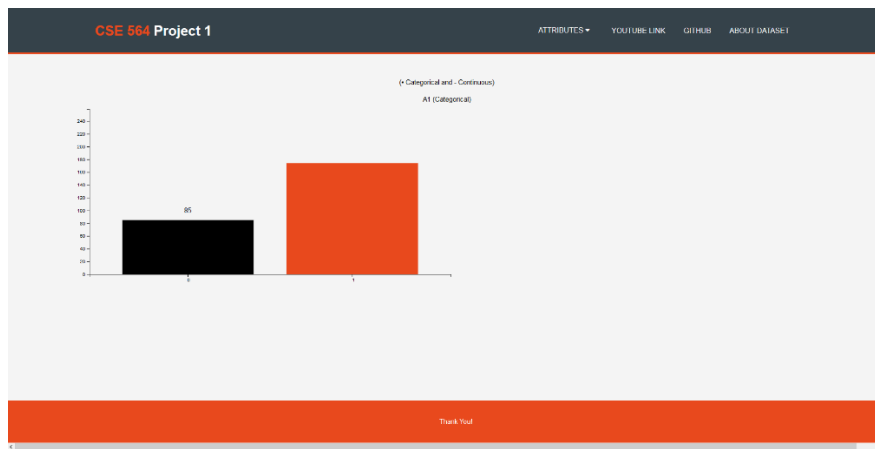
The number of values in a category shown above the bar on mouseover.

A continuous variable visualized using a bar chart. The slider will change the number of bins as shown in the video.

The number of bins has increased for A2 (Continuous).

You can select any attribute from the drop down menu which has 2 notations for differentiating between categorical and numerical: (•) Categorical and (-) Continuous.

Code Snippets:

Function for Categorical attributes:

```
function Categorical(text1,file) {
   //text1 = 'A1 (Categorical)';
   document.getElementById("variable").innerHTML = text1;
   if(!d3.select(".slidercontainer").empty())
   {
      d3.select(".slidercontainer").remove();
   }
   Barchart(file);
}
```

I clear the slider if it's there and call the Barchart function which accesses data from the file.

Function for Continuous attributes:

```
function Continuous(text1,file) {
   //text1 = 'A2 (Continuous)';
   document.getElementById("variable").innerHTML = text1;
   // currentattr = attr1;
   if(d3.select(".slidercontainer").empty())
   {
      d3.select("#svgElement").append("div")
                    .attr("class","slidercontainer")
                    .append("input")
                       .attr("type","range")
                       .attr("min",1)
                       .attr("max",200)
                       .attr("value",15)
                       .attr("class","slider")
                       .attr("id","myRange");
   }

   Histogram(file);
}
```

Code snippet to update the graph using the silder:

```
var slider = document.getElementById("myRange"); //myRange is the slider defined in the continuous  function
     var noOfBins = slider.value;

     slider.oninput = function() {
        noOfBins = +this.value;
        update(noOfBins);
        console.log("No of bins:",noOfBins);
     }
```

Code snippet to highlight the bar and display the value:

```
        .on("mouseover", function(d,i) {
                    d3.select(this)
                    .style("fill","#000000");
```

```
      tip.html( "<span style='color:black'>" + d.Count + "</span>");
      tip.show();
   })

   .on("mouseout", function(d,i) {
      d3.select(this)
      .style("fill","#e8491d");
      tip.hide();
   })
```

The interface is extremely responsive and easy to understand with a matching color scheme.

The youtube video(1:28) showing the project demonstration is: https://www.youtube.com/watch?v=KfDELReo_M0

The Y-axis range changes dynamically and the X-axis remains constant. The labeling changes dynamically.