CSE 564
Visualization & Visual Analytics

Mini Project #1

Klaus Mueller

Computer Science Department
Stony Brook University
Get a feel for data and where to find them

Get your hands dirty with JavaScript and D3.js

Online tutorials

- JavaScript: [W3Schools.com](#)
- D3: [freeCodeCamp](#), [D3 website](#), [github](#) [makeBarChart](#) [ScottMurray](#) [exampleHub](#) [dashingd3js](#) [tutorialsteacher](#)

Take advantage of this opportunity to learn D3 now

- you will need this later in the course
The variables → the attributes or properties we measured

The data items → the samples (observations) we obtained from the population of all instances

One data item

Rectangular Dataset

<table>
<thead>
<tr>
<th></th>
<th>Name</th>
<th>Country</th>
<th>Miles Per Gallon</th>
<th>Acceleration</th>
<th>Horsepower</th>
<th>Weight</th>
<th>Cylinders</th>
<th>Year</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Volkswagen Rabbit DI</td>
<td>Germany</td>
<td>43.1</td>
<td>21.5</td>
<td>48</td>
<td>1985</td>
<td>4</td>
<td>78</td>
<td>2400</td>
</tr>
<tr>
<td>3</td>
<td>Ford Fiesta</td>
<td>Germany</td>
<td>36.1</td>
<td>14.4</td>
<td>66</td>
<td>1800</td>
<td>4</td>
<td>78</td>
<td>1900</td>
</tr>
<tr>
<td>4</td>
<td>Mazda GLC Deluxe</td>
<td>Japan</td>
<td>32.8</td>
<td>19.4</td>
<td>52</td>
<td>1985</td>
<td>4</td>
<td>78</td>
<td>2200</td>
</tr>
<tr>
<td>5</td>
<td>Datsun B210 GX</td>
<td>Japan</td>
<td>39.4</td>
<td>18.6</td>
<td>70</td>
<td>2070</td>
<td>4</td>
<td>78</td>
<td>2725</td>
</tr>
<tr>
<td>6</td>
<td>Honda Civic CVCC</td>
<td>Japan</td>
<td>36.1</td>
<td>16.4</td>
<td>60</td>
<td>1800</td>
<td>4</td>
<td>78</td>
<td>2250</td>
</tr>
<tr>
<td>7</td>
<td>Oldsmobile Cutlass</td>
<td>USA</td>
<td>19.9</td>
<td>15.5</td>
<td>110</td>
<td>3365</td>
<td>8</td>
<td>78</td>
<td>3300</td>
</tr>
<tr>
<td>8</td>
<td>Dodge Diplomat</td>
<td>USA</td>
<td>19.4</td>
<td>13.2</td>
<td>140</td>
<td>3735</td>
<td>8</td>
<td>78</td>
<td>3125</td>
</tr>
<tr>
<td>9</td>
<td>Mercury Monarch</td>
<td>USA</td>
<td>20.2</td>
<td>12.8</td>
<td>139</td>
<td>3570</td>
<td>8</td>
<td>78</td>
<td>2850</td>
</tr>
<tr>
<td>10</td>
<td>Pontiac Phoenix</td>
<td>USA</td>
<td>19.2</td>
<td>19.2</td>
<td>105</td>
<td>3535</td>
<td>6</td>
<td>78</td>
<td>2800</td>
</tr>
<tr>
<td>11</td>
<td>Chevrolet Malibu</td>
<td>USA</td>
<td>20.5</td>
<td>18.2</td>
<td>95</td>
<td>3155</td>
<td>6</td>
<td>78</td>
<td>3275</td>
</tr>
<tr>
<td>12</td>
<td>Ford Fairmont A</td>
<td>USA</td>
<td>20.2</td>
<td>15.8</td>
<td>85</td>
<td>2965</td>
<td>6</td>
<td>78</td>
<td>2375</td>
</tr>
<tr>
<td>13</td>
<td>Ford Fairmont M</td>
<td>USA</td>
<td>25.1</td>
<td>15.4</td>
<td>88</td>
<td>2720</td>
<td>6</td>
<td>78</td>
<td>2275</td>
</tr>
<tr>
<td>14</td>
<td>Plymouth Volare</td>
<td>USA</td>
<td>20.5</td>
<td>17.2</td>
<td>100</td>
<td>3430</td>
<td>6</td>
<td>78</td>
<td>2700</td>
</tr>
<tr>
<td>15</td>
<td>AMC Concord</td>
<td>USA</td>
<td>19.4</td>
<td>17.2</td>
<td>90</td>
<td>3210</td>
<td>6</td>
<td>78</td>
<td>2300</td>
</tr>
<tr>
<td>16</td>
<td>Buick Century</td>
<td>USA</td>
<td>20.6</td>
<td>15.8</td>
<td>105</td>
<td>3380</td>
<td>6</td>
<td>78</td>
<td>3300</td>
</tr>
<tr>
<td>17</td>
<td>Mercury Zephyr</td>
<td>USA</td>
<td>20.8</td>
<td>16.7</td>
<td>85</td>
<td>3070</td>
<td>6</td>
<td>78</td>
<td>2425</td>
</tr>
<tr>
<td>18</td>
<td>Dodge Aspen</td>
<td>USA</td>
<td>18.6</td>
<td>18.7</td>
<td>110</td>
<td>3620</td>
<td>6</td>
<td>78</td>
<td>2700</td>
</tr>
<tr>
<td>19</td>
<td>AMC Concord D1</td>
<td>USA</td>
<td>18.1</td>
<td>15.1</td>
<td>120</td>
<td>3410</td>
<td>6</td>
<td>78</td>
<td>2425</td>
</tr>
<tr>
<td>20</td>
<td>Chevrolet MonteCarlo</td>
<td>USA</td>
<td>19.2</td>
<td>13.2</td>
<td>145</td>
<td>3425</td>
<td>8</td>
<td>78</td>
<td>3900</td>
</tr>
<tr>
<td>21</td>
<td>Buick RegalTurbo</td>
<td>USA</td>
<td>17.7</td>
<td>13.4</td>
<td>165</td>
<td>3445</td>
<td>8</td>
<td>78</td>
<td>4400</td>
</tr>
<tr>
<td>22</td>
<td>Ford Futura</td>
<td>Germany</td>
<td>18.1</td>
<td>11.2</td>
<td>139</td>
<td>3205</td>
<td>8</td>
<td>78</td>
<td>2525</td>
</tr>
<tr>
<td>23</td>
<td>Dodge Magnum XE</td>
<td>USA</td>
<td>17.5</td>
<td>13.7</td>
<td>140</td>
<td>4080</td>
<td>8</td>
<td>78</td>
<td>3000</td>
</tr>
<tr>
<td>24</td>
<td>Chevrolet Chevette</td>
<td>USA</td>
<td>30</td>
<td>16.5</td>
<td>68</td>
<td>2155</td>
<td>4</td>
<td>78</td>
<td>2100</td>
</tr>
</tbody>
</table>
Rectangular Dataset

Also called the *Data Matrix*

Car performance metrics

or Survey question responses

or Patient characteristics

One data item

Car models

or Survey respondents

or Patients
Some Good Sources For Data

Kaggle – lots of data for data science
NYC Open Data – all kinds of data related to NYC operations
Kaiser Foundation – numerous data related to public health
Data.gov – open data site with US government data
Forbes – site with links to data sites
Data Quest – another site with links to data sites
Quandl – mostly financial and economics data
Open Data Inception – map w/data portals around the world
World Bank – collection of global development data
UCI repository – site that has been around for a long time
Analytics Vidhya – another site with many links to data sites
Wikipedia also has lots of data in tables
Some advice

- avoid datasets where the attributes only have a few different values
  - this applies to both categorical and numerical attributes
- convert textual categories into numbers by assigning a numerical ID
- if dataset is too large, reduce samples by random selection (for now)
- if you have too many attributes keep the ones of interest
- highly recommended: fuse multiple datasets together to get a more holistic analysis with a deep explanation of the ecosystem;
  - data on crime stats of a city plus data on education and data on demographics (3 files from different sources)
  - data on houses for sale (w/ house properties) plus data on the zip code of the houses such as education, crime, distance to airport, etc.
  - use Google to search for data such as “education quality by zip code”
- produce a spreadsheet of rows (data items), attributes (columns)
  - the goal is to add more columns (attributes)
Assignment (1)

Get some CSV-based data (see course slides for good sources)
- at least 250 data points (the more the better)
- at least 15 dimensions (fuse datasets to achieve this)
- good mix of numerical and categorical variables (minimum 5 levels)

Your D3-based visual interface should be able to (10 pts each):
1. present a menu to allow users to select a variable and update chart
2. draw a bar chart if a categorical variable is selected
3. draw a histogram if a numerical variable is selected (bin it into a fixed range (equi-width) of your choice)
4. on mouse-over display the value of the bar on top of the bar
5. on mouse-over also make the bar wider and higher to focus on it
6. mouse (with left mouse button down) move left (right) should decrease (increase) bin width/size (for numerical variables only)
An additional 10 pts for elegant implementation/function

Don’t forget to

▪ label the axes (variable names)
▪ label the x-axis (bin range midpoints or category label)
▪ label the y-axis (number of items)

Due date is Tuesday, February 18 end of day

Submission on blackboard
Upload the following:

- 2-3 page report with illustrated description of your program’s capabilities and implementation detail
- Add code snippets to show how you did things
- YouTube link to a voice-narrated video file that shows all features of your software in action
- Zip file with source code

Grading

- TA will pick students at random for thorough code review sessions
- You better know your code!!!
- So, please do not just copy code beyond the D3 templates
- Or even worse, videotape someone else’s program
Aka, cheating

Discussion with your classmates (but not others) is OK

Cut and paste from any source is not OK

- any suspected activity of this kind will result in zero points
- also for the person providing the original
- two-strikes and out rule is in effect (including an academic misconduct report)
- this includes any feeble attempt to cover the tracks somehow

Stay honest and resist the temptation!