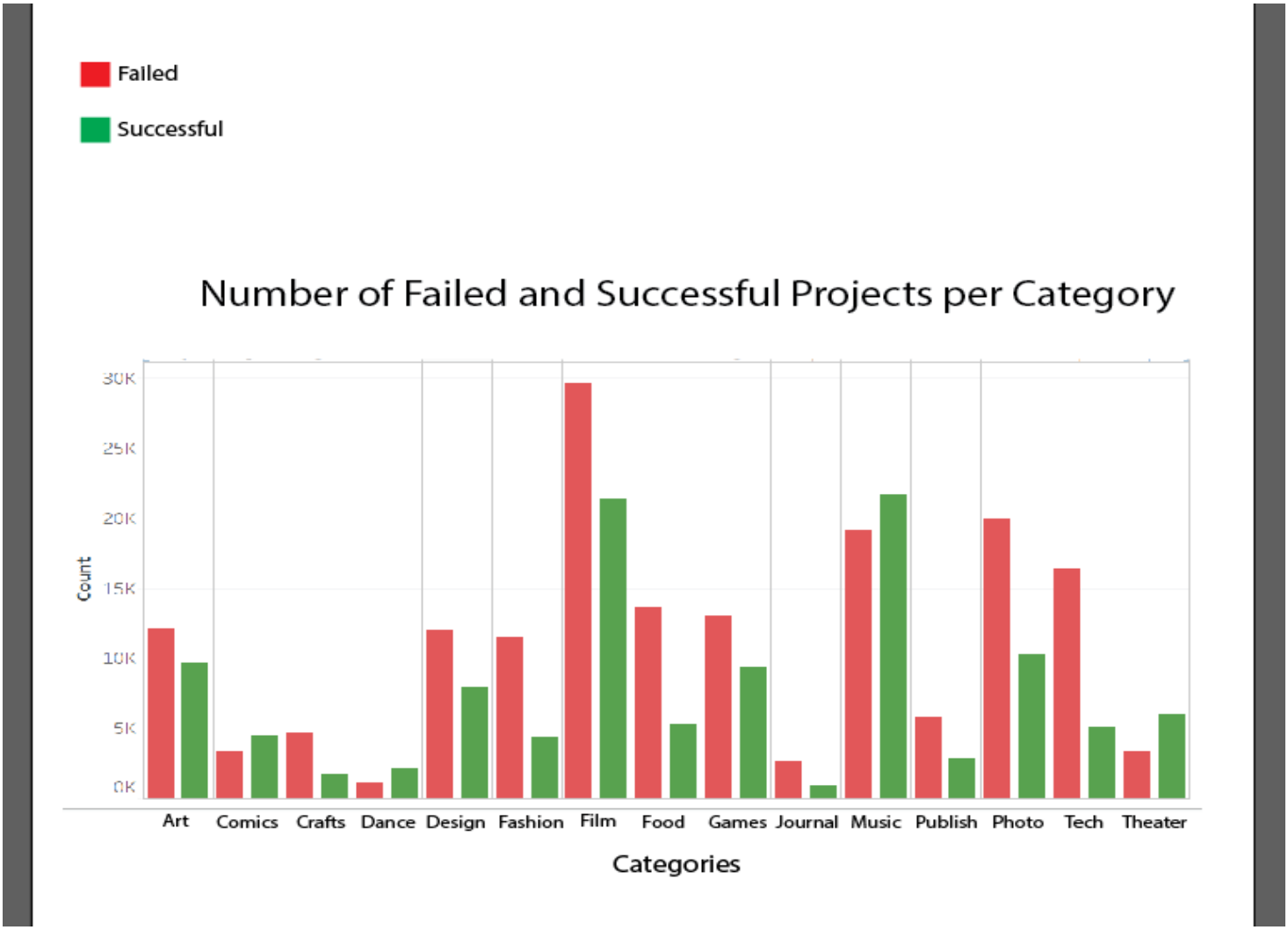
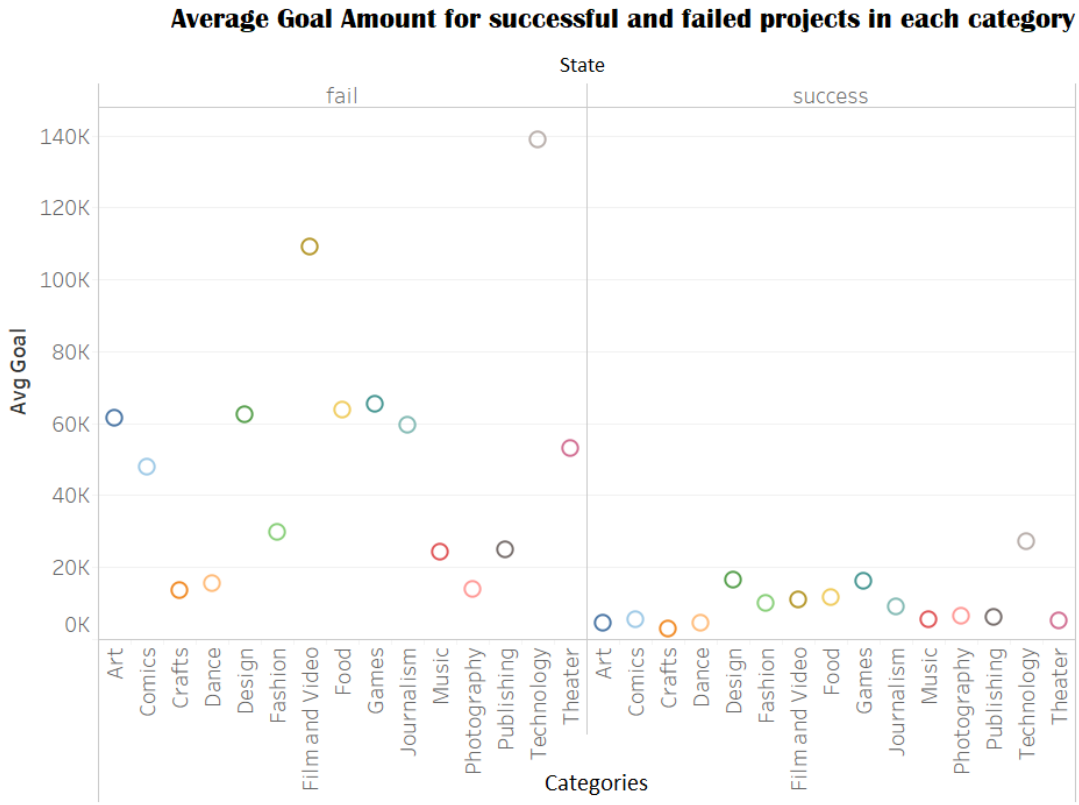
Data Visualization (SARC5400) Assignment 3 – Visual Data Analysis

The dataset I've decided to explore is the Kickstarter dataset. Kickstarter is a website that allows people to start projects they feel passionate about and raise money to execute them. They can define a 'goal' for the amount of money they wish to raise. If the goal amount is reached – the project is considered successful. The question I am trying to answer through my analysis is: "What factors affect the final state of a project?" (ie. Is the project a success or a failure?)



- This graph shows how the category of the project affects whether it is a success or not.
- The double bar chart allows the viewer to compare the number of fails and successes in each category easily
- The use of the colors: red and green have a meaning here – red is usually associated with stopping – so in this case it represents failure, as green usually means 'go' – here it represents success
- There is not a lot of information being shown in the same plot, so it is relatively easy to understand. There is no cognitive overload
- I believe this structure depicts the answer to the question quite well. The graph organizes the data by categories (column) and then state (color)

- Insights:
    - there are only a few categories with a higher number of successful projects than failed (only 4)
    - Film, music and photography have a higher number of overall projects than the other categories
    - The categories with the higher number of successful projects tend to have a lower number of overall projects
- It can be confusing to read this chart as there are only a few grid lines on the graph (technical difficulties)
- This chart does not include location or goal amount



This graph shows how the goal amount set for projects in a particular category affect whether the project succeeds or not.

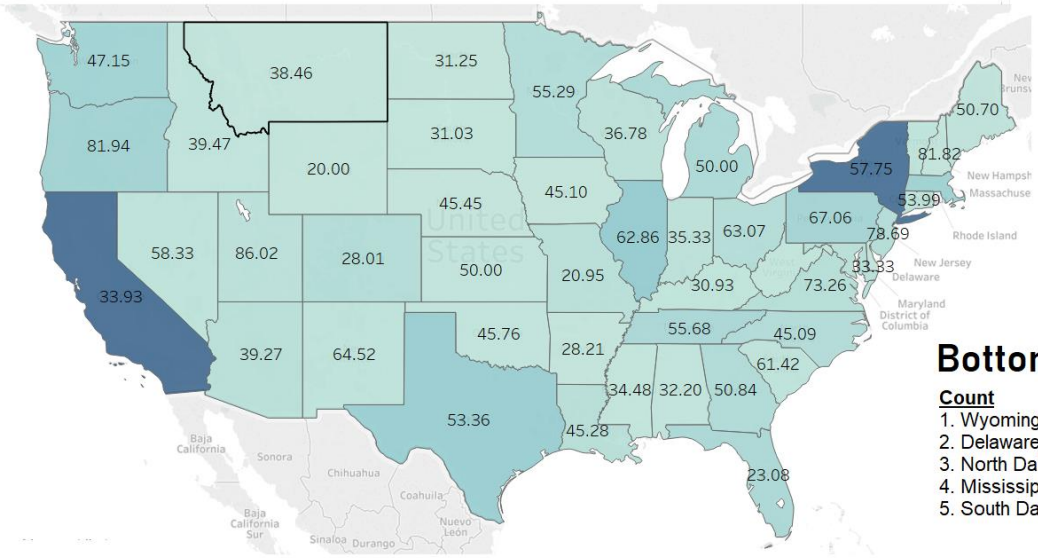
The colors allow the viewers to compare the 'failed' and 'success' points on the graphs

The side by side scatter plot attempts to offer easy comparison between the circles representing the 'failed avg goal amount' and 'success avg goal amount' – however it is not very easy to keep going back and forth between the two charts, especially because the categories are so close together – upon additional study – a stacked bar chart might have been a better choice

This chart does not show the goal amount for each project – instead it shows an average amount for each category – some of the information is lost by doing this

Also, this chart does not show location or count of projects

Percentage of Successful Projects per State



SUM(Count)
3 4,427

The gradient gives you an idea of the total count of projects from a state
(Darker blue = higher count of projects)

Top 5 States:

Count	Percentage Success
1. California	1. Utah
2. New York	2. Oregan
3. Illinois	3. New Hampshire
4. Texas	4. New Jersey
5. Maryland	5. Virginia

Bottom 5 States:

Count	Percentage Success
1. Wyoming	1. Wyoming
2. Delaware	2. Missouri
3. North Dakota	3. Florida
4. Mississippi	4. Arkansas
5. South Dakota	5. West Virginia

This chart shows the distribution of the number of projects launched (in the form of a heat map) and the percentage of successful projects in each state.

The gradient gives you an idea of the total count of projects from each state. The darker the blue – the higher the count. This gradient allows for easy understanding of project count – rather than adding another label to each state showing the number of projects launched (in which case the map would be too cluttered)

There is also a list of "top 5 states" provided both in the context of count of projects and percentage of success – it can be seen that there is no overlap between the two lists – this means that there is little to no correlation between number of projects launched to the number of successful projects in each state.

The list: "bottom 5 states" has one overlapping state – Wyoming. Otherwise, it too shows that there is little relation between the number of projects launched and the number of successful projects per state.

This chart would have been better if there was some sort of a time scale – it would have more meaning if we knew whether this data was collected over one year or since the start of the website.

This particular method of graph did not allow for the depiction of percentage of failed projects – however, it can be determined by subtracting the success percentage from 100%

It also did not include the category of the project