

**Due: 11:59pm, Feb. 9, 2024**

## Learning Objectives

The goal of this assignment is to become acquainted with tree maps and iterative design processes.

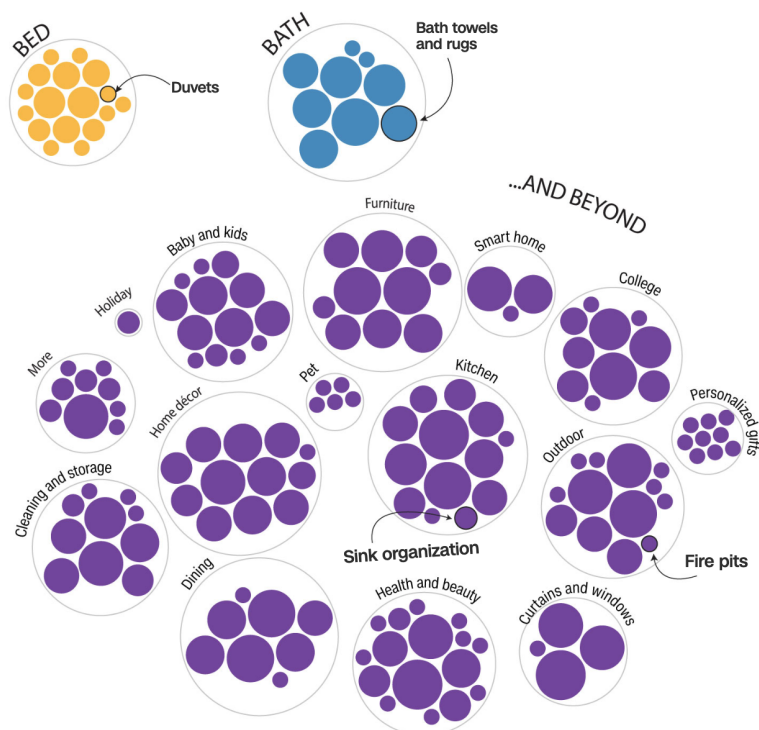
## Data

Download the `bbb.csv` data set from Brightspace.

This data contains information about products sold at the (now bankrupt) retail chain Bed, Bath & Beyond. Each top level category is broken down into several un-named items, each with some value that represents the number links on the BBB website associated with that sub-category.

This data comes from the following article, with their visualization of the data below:

- How big was the ‘beyond’ section at Bed, Bath & Beyond? The iconic retailer’s weirdest merchandise, *CNN*, <https://www.cnn.com/2023/04/26/business/bed-bath-beyond-merchandise-dg/index.html>



## Instructions

Using the provided data set, create a Jupyter notebook to answer the following questions.

You will need to use the squarify (Q1–Q3) and circlify (Q4–Q5) libraries. You may only import pandas, Matplotlib, squarify, circlify, and math or NumPy.

Squarify has 3 primary function calls that will be required: `normalize_sizes()`, `squarify()`, and `plot()`, which is a roll-up of the previous two calls plus visualization of the result. The limited documentation for squarify will be helpful.

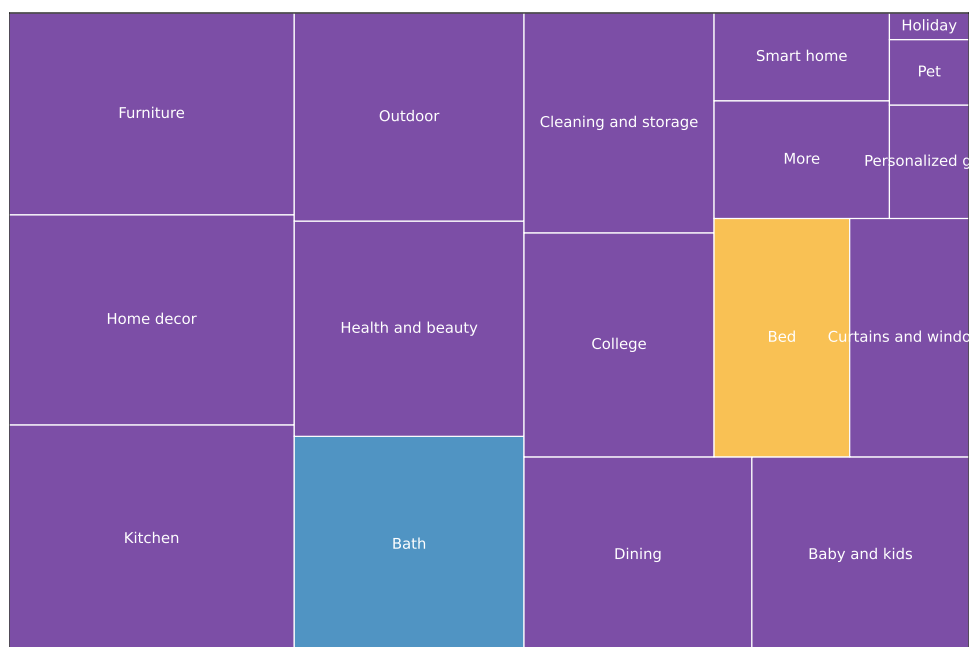
Circlify has 1 primary function – `circlify()`. I advise consulting the documentation.

The colours needed for this assignment are '#7c4ea6', '#5094c3', and '#f9c154'.

### Question 1 – Basic Tree Map: (20 pts)

A tree map is a hierarchical grouping of items into rectangular shapes. They are tiled to fill the space, and, importantly, the area of each rectangle is proportional to the quantitative value of that category.

For this question, create a valid tree map of the top level categories in the `bbb.csv` data set using Matplotlib and the squarify libraries, like the example below. The area of each rectangle is scaled to the total value of each category. For example, the total value of the 'Kitchen' category is 1063.7, whereas 'Pet' is 85.1.



Q1: FIRST DESIGN

In addition to being a valid tree map, your design should:

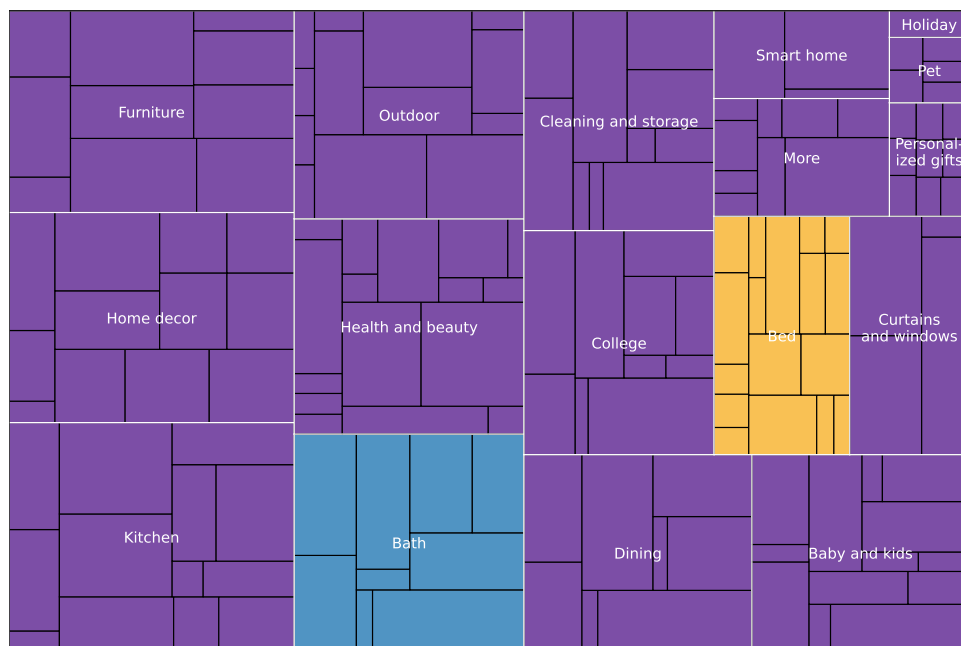
- label each rectangle with its particular category in the centre of the rectangle,
- colour each rectangle as shown above (reflecting the colour scheme of the original image),
- remove any unnecessary tick marks or labels, and
- delineate each rectangle by white lines.

The tiling of your rectangles may be different than the example above. That is ok, as long as it is a valid tree map. The outcome of your tiling is dependent upon a number of factors, in particular, the aspect ratio of your image.

## Question 2 – Adding Hierarchy: (30 pts)

The tree map in Question 1 lacks hierarchy and has several deficient elements. Each category (Bed, Bath, etc) is made up of a number of items, which were not represented. In addition, some text labels extend outside of their rectangle or out of the frame of the visualization.

For this question, extend the tree map from Question 1 to include the items within each category, like the example below. All text labels should be word wrapped to stay within the boundaries of its respective rectangle.



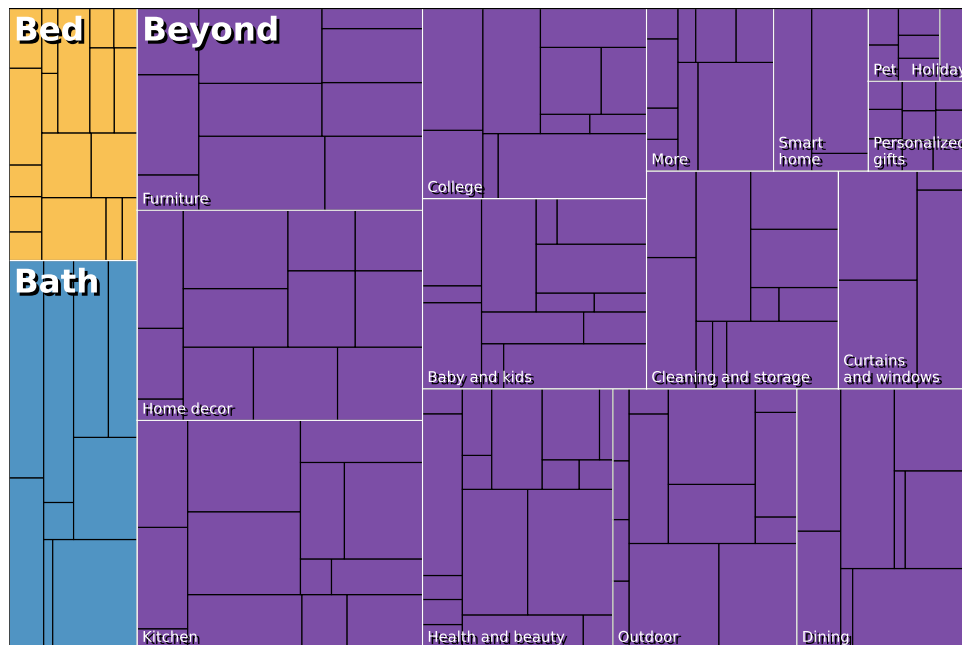
Q2: LAYERING HIERARCHY

Importantly, the areas of all rectangles should be proportional to its corresponding quantitative value. The area of each top level rectangle represents the sum of its constituent rectangles. Losing this consistency would distort the meaning of the visualization and cause misunderstanding.

### Question 3 – Finishing Touches: (20 pts)

Let's improve our tree map from Question 2 in two further ways. Firstly, the 'Bed' and 'Bath' categories are intermixed with the other categories. This doesn't make sense. All of the purple rectangles represent a 'Beyond' category and should be placed into a single large rectangle. Let's fix that.

Secondly, the text is difficult to read. Let's reinforce our top-level labelling of Bed, Bath and Beyond, and add a shadowing to the text to help make it visually distinct.



Q3: FINISHING TOUCHES

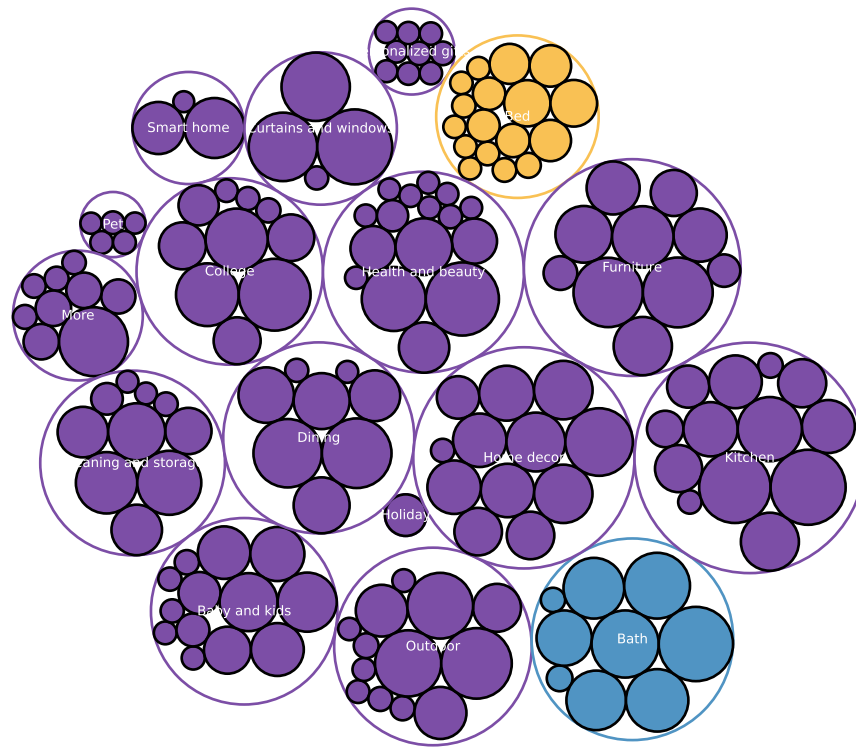
Importantly, again, the areas of all rectangles should be proportional to its corresponding quantitative value. This has not changed from previous versions. Maintaining data integrity is a fundamental requirement of data visualization.

The exact tiling may differ from the example above. However, it should consist of 3 large rectangles, one for Bed, one for Bath and one for Beyond. The Beyond rectangle should be sub-divided into 15 rectangles, one for each category. These are further subdivided by their sub-categories.

All text elements should follow the example above. The 3 top level categories should be a large font size, in bold, and placed in the top left of their respective rectangle. All text should include shadowing. Items within the 'Beyond' rectangle should have their labels in the bottom left. Text should be word wrapped to fit inside its corresponding rectangle (where possible; in my example 'Holiday' cannot fit inside the rectangle, but does fit within the frame of the visualization).

### **6934 Students Only – Question 4 - Packed Circle Chart: (15 pts)**

Use Matplotlib and the circlify package to create a packed circle chart of the `bbb.csv` data, like in the example below.

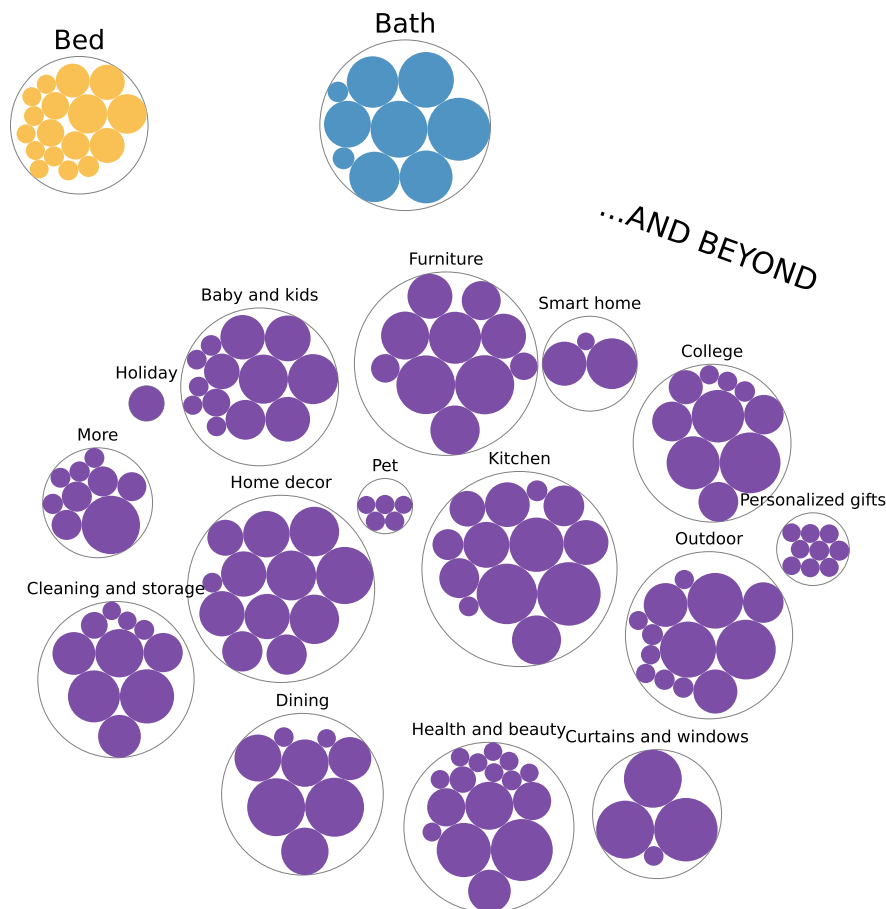


As with the tree map, the area of each circle is proportional to the quantitative value associated with that item. Items within a category are arranged to pack tightly within an inscribing circle. In a sense, this is like a tree map but built using circles instead of rectangles. (Though representing the hierarchy can be difficult as the interior may not maximally fill their inscribing circle.)

Make sure that the colours of the items within a category and the enclosing circles are correct, and the category is labelled in the middle of the enclosing circle (we will improve this on the next iteration).

## 6934 Students Only – Question 5 - Final Version: (15 pts)

Adjust your packed circle chart from Question 4 to look like the original image from the CNN article. An example is below.



The exact placement of items within each category do not need to be exactly as in the original image, but the categories themselves (enclosing circles) should be positioned correctly. Elements that should be present in your final submission include:

- each category is positioned like the original image,
- the items within a category do not have an edge colour,
- the enclosing circle for each category is grey in colour,
- the label for each category is above its enclosing circle, and
- the text ‘...AND BEYOND’ is present, rotated 20 degrees, and positioned correctly.

Bonus marks if the text is curved around each circle like in the original image. (The four items that are specifically annotated in the original image, e.g., ‘Fire pits’, are not included in the data set, so no bonus marks for adding them.)

## Submission

Submit your Jupyter notebook (.ipynb) through Brightspace.

Late submissions will be subject to a 10% penalty for each hour past the deadline.

## Attribution

Submissions should include an attribution section indicating any sources of material, ideas or contribution of others to the submission.

Submissions must represent your independent work.

You are encouraged to use any resources to help with your solution, but your solution must represent independent work. If your submitted work includes unacknowledged collaboration, code materials, ideas or other elements that are not your original work, it may be considered plagiarism or some other form of cheating under MUN general regulations 6.12.4.2 (4.12.4.2 for graduate students) and academic penalties will be applied accordingly.

Avoid academic penalties by properly attributing any contribution to your submission by others, including internet sources and classmates. This will also help distinguish what elements of the submission are original. You may not receive full credit if your original elements are insufficient, but you can avoid penalties for plagiarism or copying if you acknowledge your sources.

## Github

I encourage you to store and version your work on GitHub. It is good practice to do so as everyone uses git in the real world.

However, **it is a requirement that git repositories containing assignment material be private.** University regulations (undergraduate 6.12.4.2 and graduate 4.12.4.2) consider it cheating if you allow your work to be copied. There will be zero tolerance for this.