CS 573: Final Project

PROCESS BOOK

Özge Aygül

Dynamic Course Scheduling for Strategic University Scaling

via Mathematical Optimization

(LINK)

This project aims to create a scrollytelling website that helps nonexpert stakeholders to understand how mathematical optimization is useful for university scheduling, as well as strategic planning.

Overview & Motivation

Project Goals: Elaborate on the aim to make mathematical optimization accessible and understandable through scrollytelling, focusing on its application in university scheduling and strategic planning.

Motivation: Discuss the challenge non-expert stakeholders face in understanding mathematical optimization and how this affects their decision-making and involvement in university operations.

Related Work

- **Classroom Reflection Sessions:** Throughout the course, reflection sessions have been a source of inspiration for my project. These sessions often featured a variety of scrollytelling visualizations that demonstrated effective ways of communicating complex information through engaging, narrative-driven formats. Observing these examples helped me understand the power of storytelling in data visualization and guided my approach to making mathematical optimization accessible to non-experts

- **Challenges in Field Communication:** In the field of mathematical optimization, one of the primary challenges is the translation of optimal solutions into strategies that are comprehensible and actionable for stakeholders who may not have a technical background. This issue is especially prevalent in university settings, where decisions based on optimization techniques can have significant impacts on resource allocation and strategic planning. By using scrollytelling to visualize optimization processes and results, my goal is to bridge the gap between complex mathematical theories and practical, real-world applications.

Questions

- **Initial Questions**
- How can mathematical optimization techniques improve university scheduling?
- What are the typical challenges faced by university administrators that these techniques can address?
- How can the benefits of these solutions be effectively communicated to non-expert stakeholders?
- **Evolution of Questions**
- What are the most impactful examples of optimization in university scheduling that resonate with non-experts?
- How can interactivity enhance understanding and engagement with complex optimization processes?
- This evolution reflects a shift from a broad exploration of optimization to a more targeted investigation of its most effective applications and communicative strategies.

Questions

New Questions
Throughout the development process, particularly during the design and implementation phases, new questions emerged that further shaped the project.

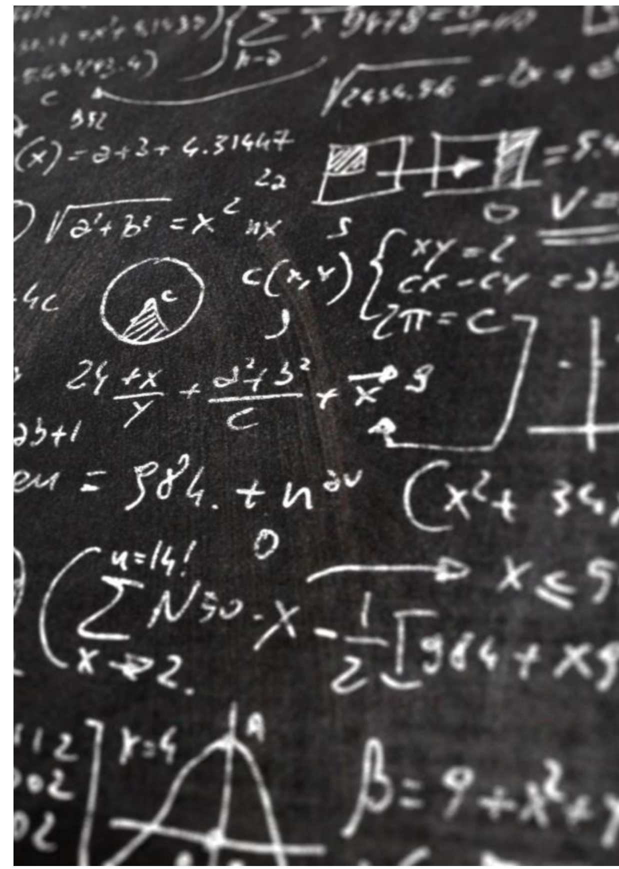
•What visualization techniques are most effective in simplifying the concept of optimization for a ***non-technical audience***?

•How can we measure the impact of our scrollytelling approach on the audience's understanding and ***perception of mathematical optimization***?

•Are there additional stakeholder ***concerns or misconceptions*** that we need to address through our visualizations?

Data

- **Actual Assignments:** This data is given by WPI for a funded research project in 2022. This data includes 2022 Fall A course room and time pattern assignments, and room capacities.
- **Optimized:** This data is generated by a mathematical optimization algorithm, developed by myself and my advisor Professor Andrew Trapp. It shows *best* assignments with respect to a variety of constraints, while considering multiple objectives.
- **Long term scenario:** This analysis answers what-if scenarios based on different university scaling plans. It shows the tradeoff between additional students in 4 years and required number of sections required to accommodate those students.

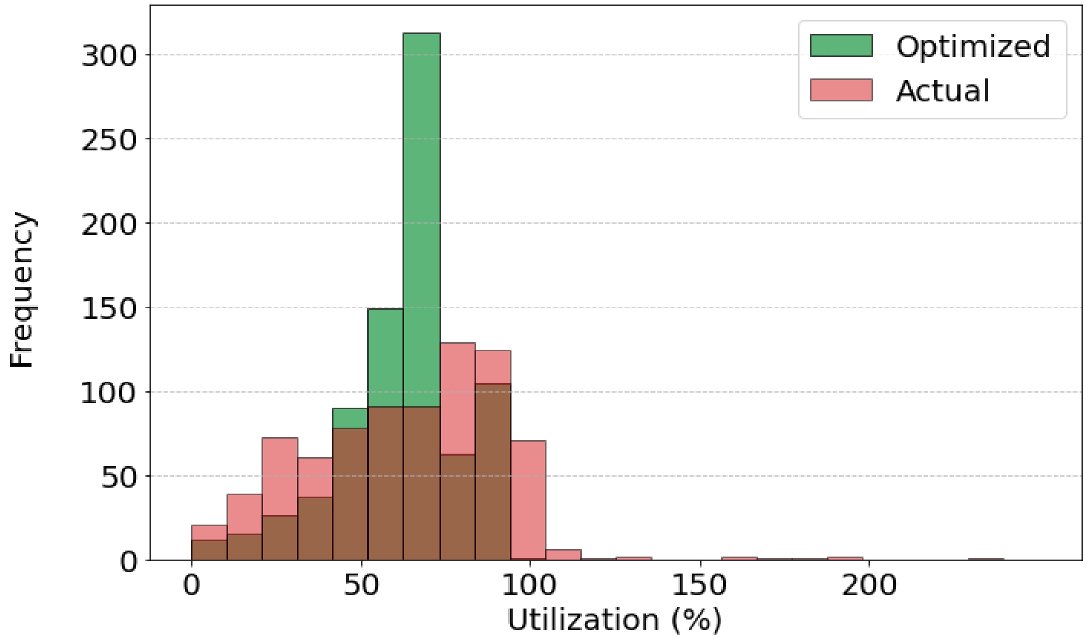


Exploratory Data Analysis

- **Initial Visualizations**

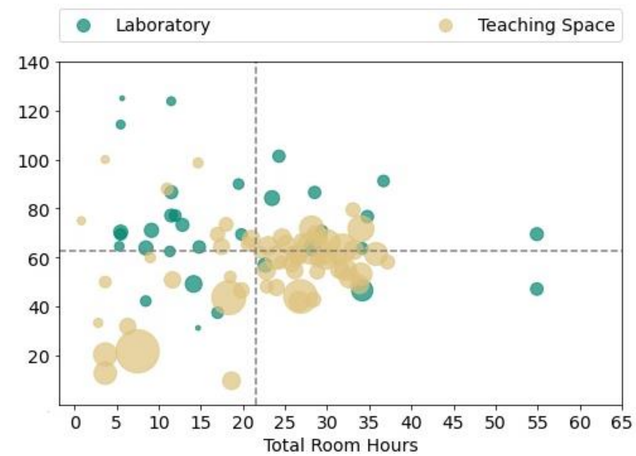
Despite my familiarity with the, I began the exploratory data analysis with a series of basic visualizations to remind myself with its characteristics and to identify the most effective ways to communicate these concepts visually. These initial visualizations included:

- **Histograms and bar charts** to show the distribution of classroom usage in the actual and optimized schedule.

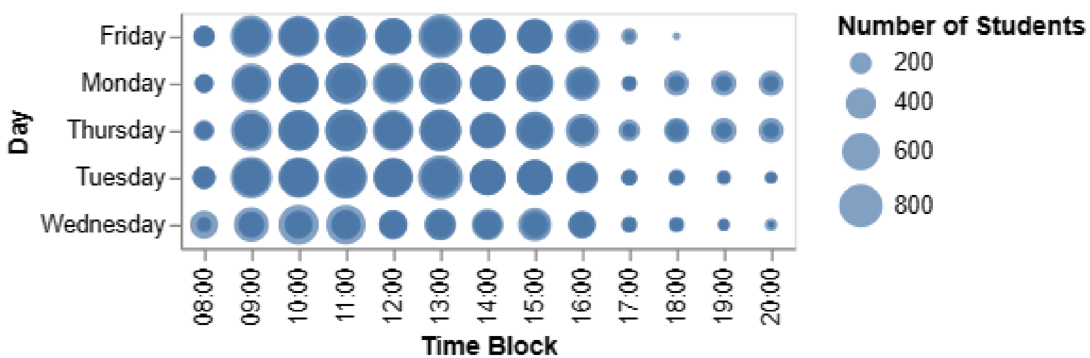


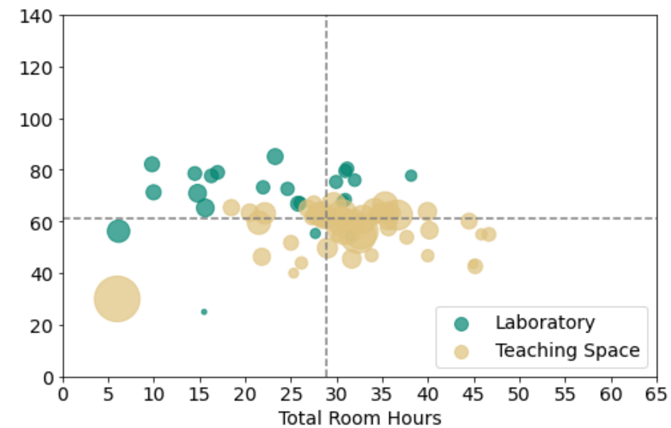
Exploratory Data Analysis

- **Scatter plots** for location utilization and usage frequency



- **Bubble chart** to show the student distribution across a typical week.





Exploratory Data Analysis

Insights

- Comparison charts shows significant improvement when optimization algorithm is used.
- Interactively show these impact is a great way to convince the audience.
- Student distribution is not the focus of the model, but it can be used to estimate the number of students in the campus at a given time to manage operations (shuttle, cafeteria, cleaning) better.
- Adding a new dimension regarding buildings can be useful.

Design Evolution & Implementation

#1: Scrollytelling for mathematical model

I want to keep the website user friendly, and easy to follow with mathematical expressions.
Therefore, I first think of embedding mathematical formulations inside a paragraph.
But it is not what I wanted, so I use idyll to go through mathematical formulations.

Step 1:

Think of mathematical optimization like a really smart way to make the best choices possible.

Imagine you've got a bunch of options—those are your **decision variables**. They're the things you can tweak to change the outcome.

$$x_s^{p\ell}, z_s^{p\ell}, q_s^{p\ell}, w_s, y_\ell, t_s, \theta_s^\ell$$

Design Evolution & Implementation

#1: Scrollytelling for mathematical model

Step 2:

$$\sum_{p\in\mathcal{P}_s}\sum_{\ell\in\mathcal{L}_s} x_s^{p\ell} = w_s \quad \forall\, s \in \mathcal{S}\setminus\hat{\mathcal{S}}, \tag{1a}$$

$$\sum_{p\in\mathcal{P}_s}\sum_{\ell\in\mathcal{L}_s} (t_s x_s^{p\ell} + z_s^{p\ell}) = t_s w_s, \quad \forall\, s \in \hat{\mathcal{S}}, \tag{1b}$$

Now, you can't just do whatever you want; you've got some rules to follow—those are your **system of constraints**.

They keep everything in check and make sure you're playing fair.

(1c)

(1d)

$$\lambda_\ell + \theta_s^\ell \le \mu_\ell, \quad \forall\, s \in \mathcal{S},\ \ell \in \mathcal{L}_s, \tag{1e}$$

$$\sum_{s\in\{\mathcal{S}_\ell\cap\mathcal{S}_k\}}\ \sum_{p\in\mathcal{P}_s^{(d,h)}} x_s^{p\ell} + \sum_{s\in\{\mathcal{S}_\ell\cap\mathcal{S}_k\cap\hat{\mathcal{S}}\}}\ \sum_{p\in\mathcal{P}_s^{(d,h)}} z_s^{p\ell} \le y_\ell,$$

$$\forall\, \ell \in \mathcal{L},\ k \in \mathcal{K},\ (d,h) \in \mathcal{DH}, \tag{1f}$$

$$\sum_{s\in\{\mathcal{S}_i\cap\mathcal{S}_k\}}\ \sum_{p\in\mathcal{P}_s^{(d,h)}}\ \sum_{\ell\in\mathcal{L}_s} x_s^{p\ell} + \sum_{s\in\{\mathcal{S}_i\cap\mathcal{S}_k\cap\hat{\mathcal{S}}\}}\ \sum_{p\in\mathcal{P}_s^{(d,h)}}\ \sum_{\ell\in\mathcal{L}_s} z_s^{p\ell} \le 1,$$

$$\forall\, i \in \mathcal{I},\ k \in \mathcal{K},\ (d,h) \in \mathcal{DH}_i, \tag{1g}$$

Here, I centered the constraints as they are wider.

The goal is to show there is a complex math behind; not go through every constraint.

Design Evolution & Implementation

#1: Scrollytelling for mathematical model

Steps 4-5-6-7:

The first objective function aims to maximize the number of course sections assigned, while minimizing capacity relaxation and the number of splits.

The values of ρ_1 and ρ_2 are intentionally set to create a preferred sequence of priorities in an ordered fashion

$$maximize \quad f_1(\cdot) = \sum_{s \in \mathcal{S}} w_s - \rho_1 \sum_{s \in \hat{\mathcal{S}}} \sum_{\ell \in \mathcal{L}_s} \theta_s^\ell - \rho_2 \sum_{s \in \hat{\mathcal{S}}} t_s.$$

The second objective function aims to maximize the (soft) preferences of instructors.

$$maximize \quad f_2(\cdot) = \sum_{s \in \mathcal{S}} \sum_{p \in \mathcal{P}_s} \sum_{\ell \in \mathcal{L}_s} v_s^{p\ell} x_s^{p\ell} + \sum_{s \in \hat{\mathcal{S}}} \sum_{p \in \mathcal{P}_s} \sum_{\ell \in \mathcal{L}_s} v_s^{p\ell} z_s^{p\ell}.$$

Third is to minimize the total number of teaching and lab spaces to improve locational efficiency by fitting more course sections into the space made available.

$$minimize \quad f_3(\cdot) = \sum_{\ell \in \mathcal{L}} y_\ell.$$

The fourth objective is to optimize the assignments in a manner that minimizes unused capacity in each location, seeking efficient utilization of all used resources.

$$minimize \quad f_4(\cdot) = \sum_{s \in \mathcal{S}} \sum_{p \in \mathcal{P}_s} \sum_{\ell \in \mathcal{L}_s} \left(\lambda_\ell c_\ell - r_s^n\right) x_s^{p\ell} + \sum_{s \in \hat{\mathcal{S}}} \sum_{p \in \mathcal{P}_s} \sum_{\ell \in \mathcal{L}_s} \left(\lambda_\ell c_\ell - q_s^{p\ell}\right) z_s^{p\ell}.$$

In these steps, I introduce objective functions along with the mathematical expressions.

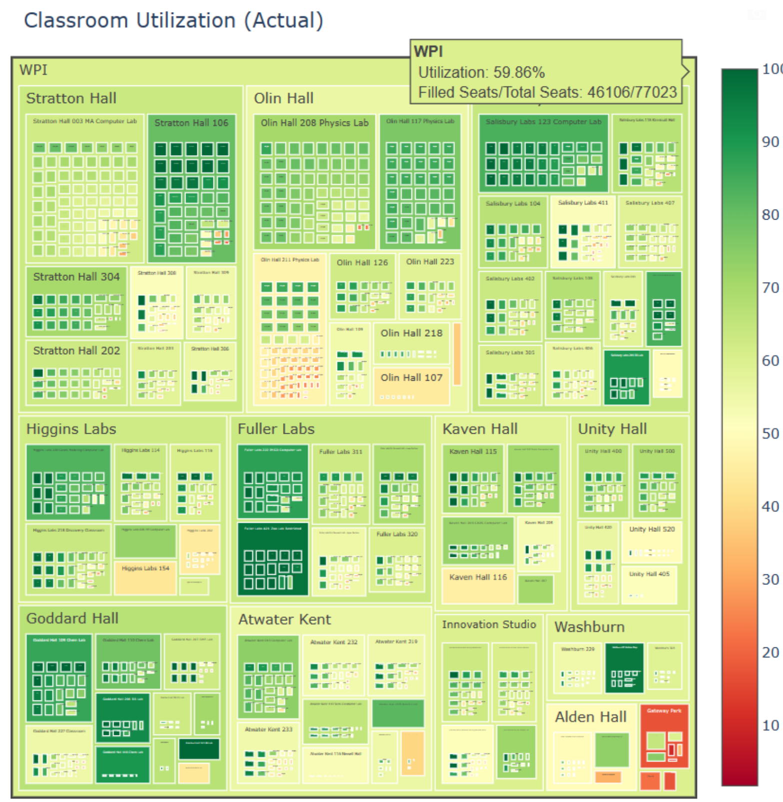
Again, the goal is to make audience familiar with the concept.

Each mathematical expression is changed seamlessly as user navigates the next steps.

Design Evolution & Implementation

#2: Treemaps ([actual](), [optimized]())

For the treemap visualizations on classroom utilization, I considered interactive elements essential for conveying the complex data in a user-friendly manner. The design decision to employ a gradient color scale from green to red allows for quick visual cues on room occupancy levels, adhering to the pre-attentive processing principle where color differences can be immediately detected by the visual system. The use of squares within larger building blocks facilitates an easy comparison between individual rooms and buildings, applying the Gestalt principles of proximity and similarity to organize information efficiently.

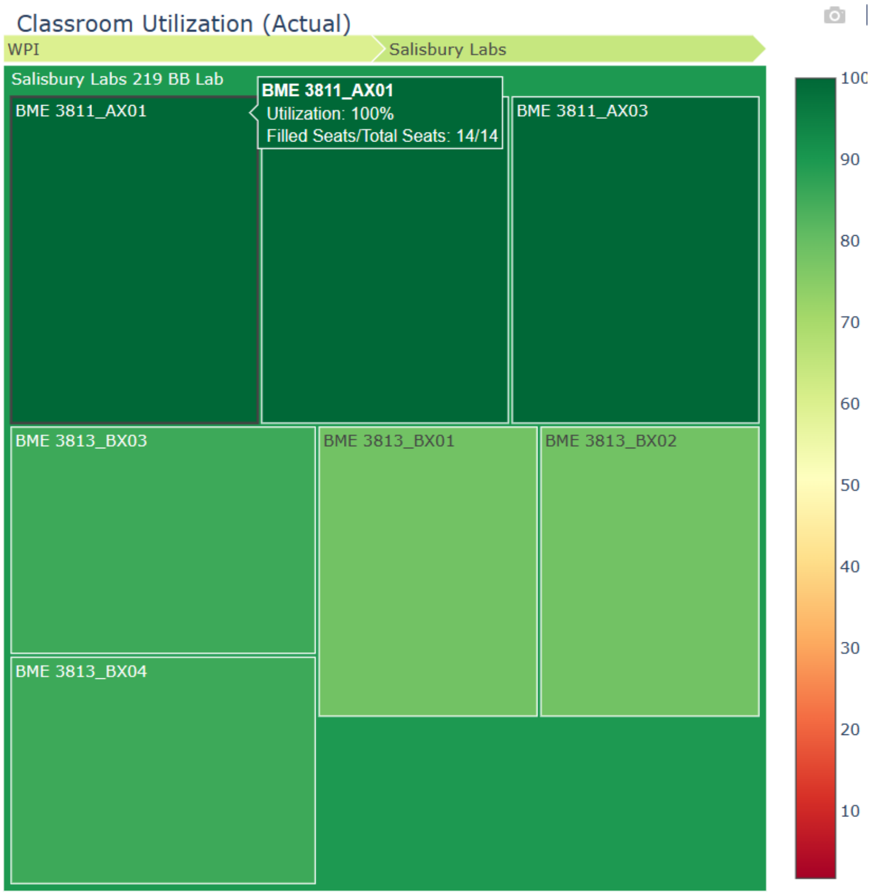


Design Evolution & Implementation

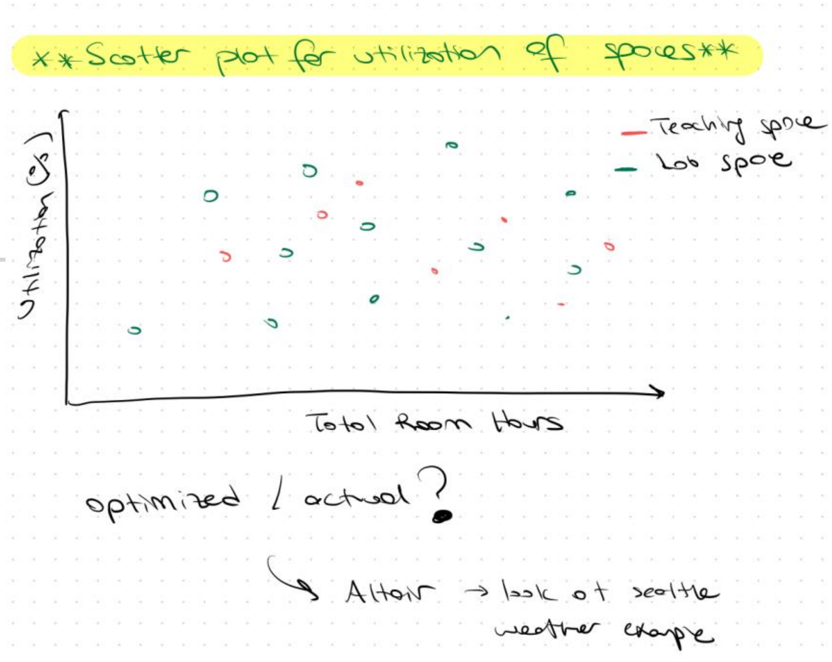
#2: Treemaps (actual, optimized)

I designed treemaps to minimize cognitive load by presenting a high-level overview, with details available upon interaction, which prevents information overload and aligns with the principle of information layering.

This design also supports the principle of focusing the viewer's attention on the most critical data first, with the option to delve deeper if needed.



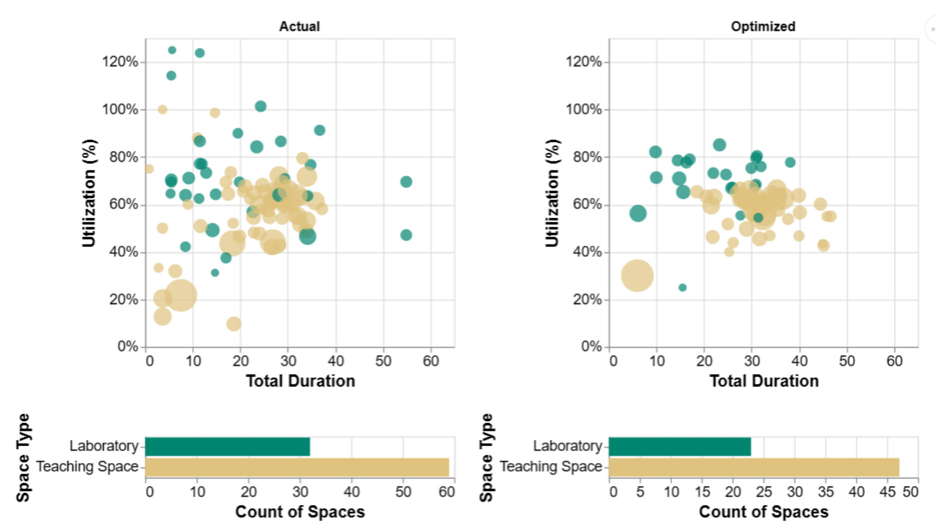
Design Evolution & Implementation



#3: Scatter Plot ([link])

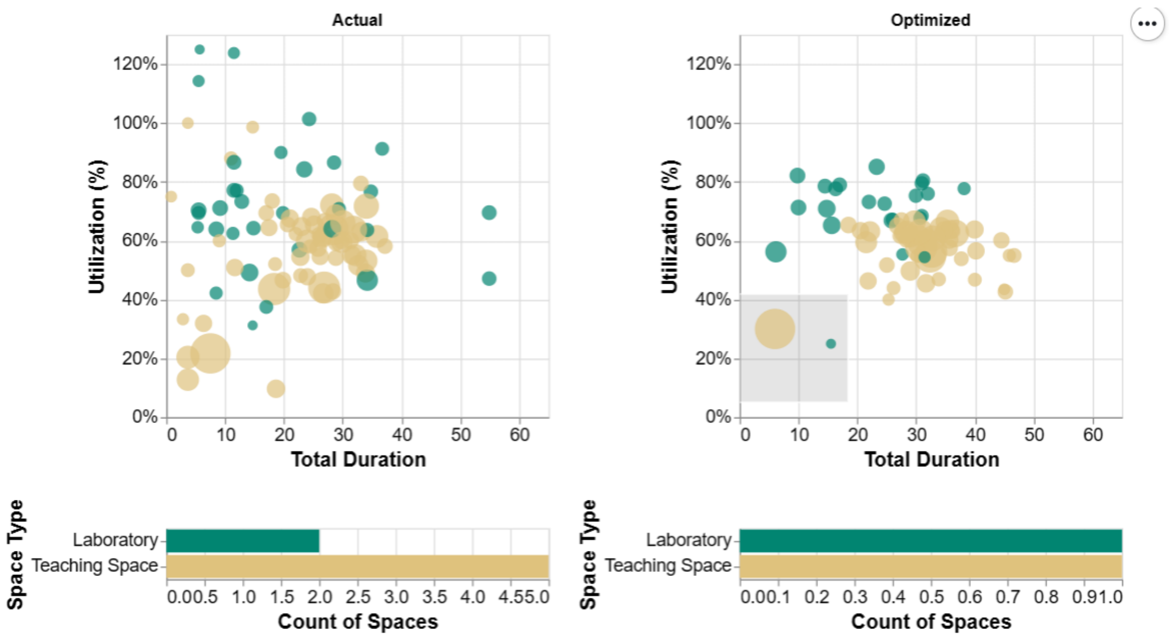
In analyzing the interactive scatterplot, I chose a dual-view approach to compare 'Actual' versus 'Optimized' utilization of space.

This visualization allows stakeholders to actively engage with the data, selecting specific areas of the plot to understand the distribution and characteristics of space utilization more deeply. The design utilizes bubble size to represent the capacity of spaces, effectively applying data encoding principles to translate quantitative information into visual form.



Design Evolution & Implementation

#3: Scatter Plot (link)



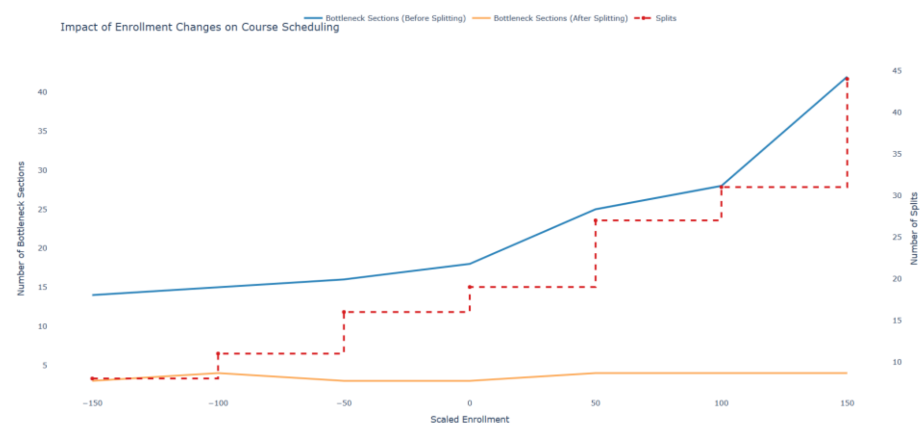
The bottom-left quadrant is particularly insightful as it highlights underutilized spaces. By interacting with this area in both 'Actual' and 'Optimized' plots, users can quickly assess the effectiveness of optimization strategies. The juxtaposition of the two states underscores the improvements and reveals areas for potential further optimization.

Design Evolution & Implementation

#3: Tradeoff Scrollytelling visualization

In this visualization, the idea is to show how number of additional student will impact the schedule. My first idea is to show how number of bottleneck sections before and after splitting algorithm is applied, but then I seemed too complicated.

Design



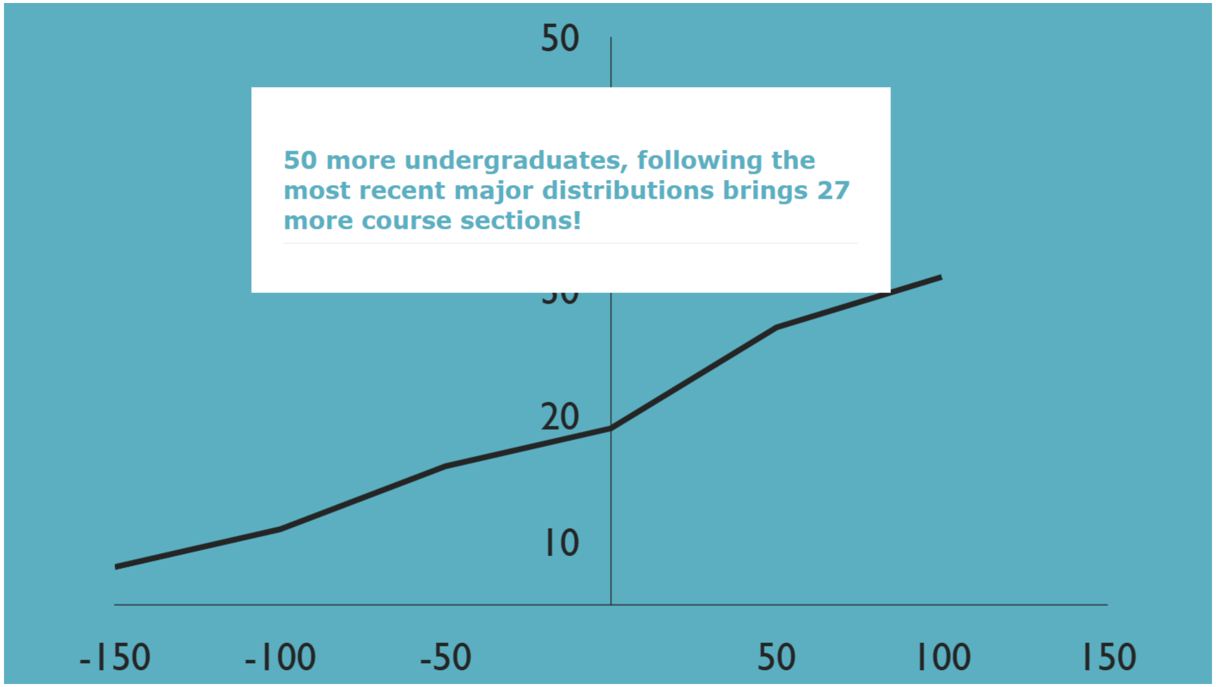
First Implementation (link)

Design Evolution & Implementation

#4: Tradeoff Scrollytelling visualization

Therefore, I decided to show only increase in the number of additional course sections required (y axis) to accommodate more students (x axis).

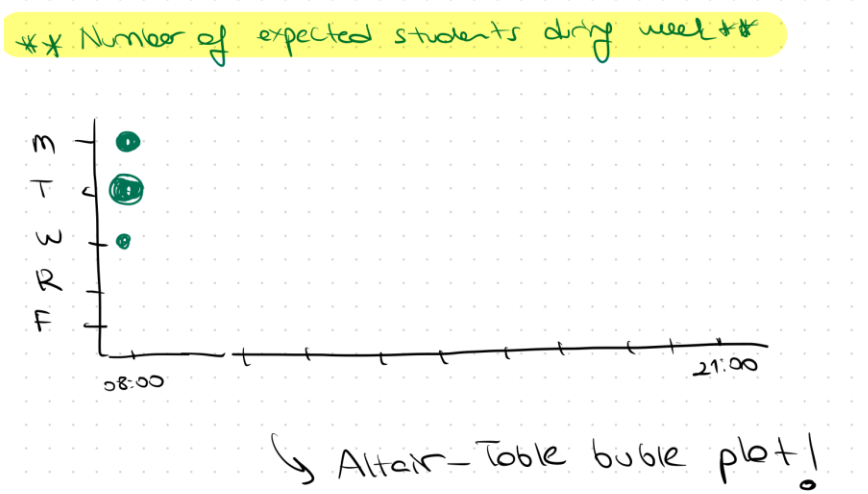
I implemented this basic line chart with idyll chart components. As user scroll down, chart moves along x axis, and information box pops up to inform the user.

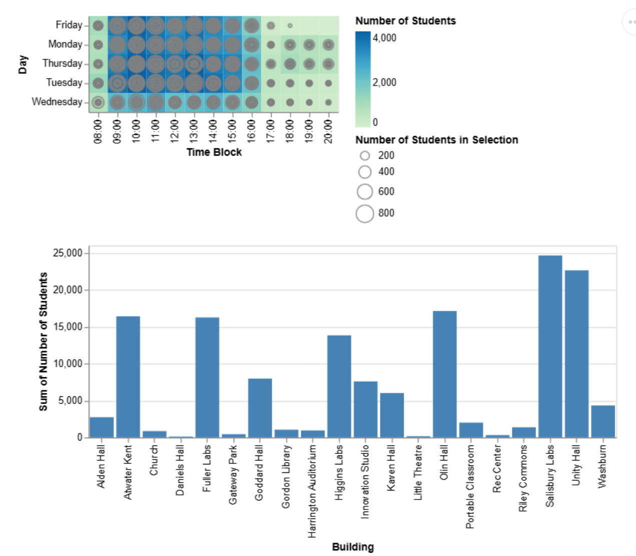


Design Evolution & Implementation

#5: Additional Charts - Heatmap

I thought it would be a good idea to show the distribution of students across buildings in a day. As I mentioned earlier, this was not a part of the optimization algorithm. However, it could help to manage operations more efficiently.



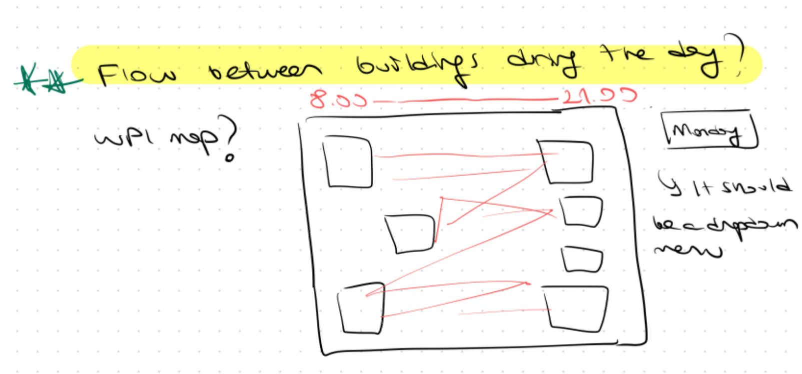


The initial design considers overall student count in the campus. However, I wanted to add building dimension. It's an effective way to visualize time-based patterns in student attendance and allows stakeholders to identify peak times and days with higher footfall.

This interactive chart (link) shows the heatmap at the campus. The bar chart shows the sum number of students across various buildings, providing a clear visual hierarchy that immediately indicates which buildings have the highest attendance. This can help in strategic planning for resource allocation and utilization. However, I did not include this visualization, as I think it did not fit to my story.

Design Evolution & Implementation

#6: Additional Charts – Moving Bubble Chart

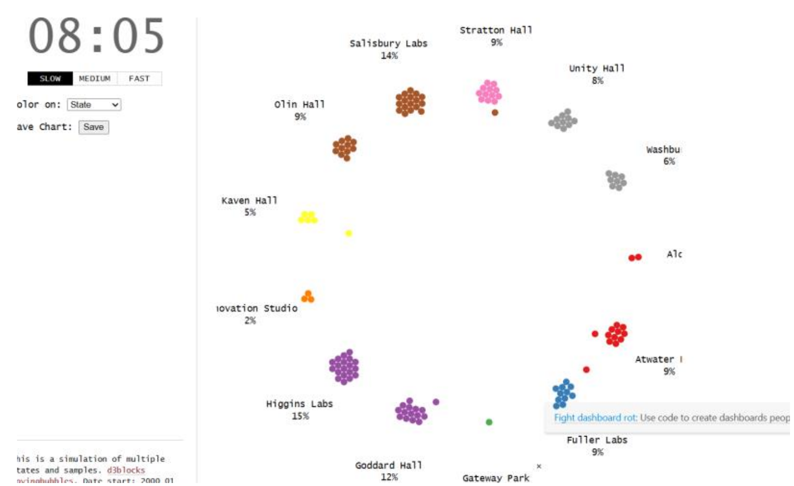


Another idea is to show the flow between buildings in a typical day. I tried to show it for Monday, first.

I implemented it with d3blocks (link).

This is again not a part of the optimization model, but there are models to aim the minimize the flow within the campus, especially campus is large in size.

I have not included this chart in my website, as I think there are issues in representing the flow.



Evaluation

- The process of creating my first scrollytelling website taught me the significance of interactive visualizations in storytelling. By embedding interactivity, like selecting time blocks to see student counts, users could actively engage with the data, uncovering insights about campus utilization. However, as a first-time web designer, I recognized the need to streamline the user journey further. To improve, I could offer more contextual tooltips, simplify navigation, and implement responsive design features to enhance accessibility and ease of use on various devices. It's about fine-tuning the balance between data exploration and guided discovery to create an even more immersive educational experience.

- Although my visualizations effectively communicated complex concepts and facilitated engagement, they could be improved by incorporating user feedback to refine interactivity and clarity. For example, enhancing the legend and instructions for interactivity might help users more intuitively understand the data, especially for those less familiar with such analysis. Also, accessibility could be further improved, ensuring that the visualizations are inclusive for all audiences, including those with visual impairments.

- The visualizations worked well to answer my core questions regarding the application of optimization in university scheduling and the challenges faced by administrators. They also succeeded in making a technical subject accessible to non-experts. However, continued user testing could reveal further opportunities for refinement.