



Student Process Visualization



Andrew Olin

Jonathan Denning

Introduction

We studied how students worked on computer science course projects. Each project provided the students with starter source code, criteria and constraints to bear in mind when approaching the project, and maybe a hint or two along the way. Beyond that, the students have free reign; this freedom allows the student to display their problem solving skills as well as their understanding of the course content thus far.

Data

Workflow data (Tbl. 1) was collected from twenty-five COS 265 Data Structures and Algorithms students from two different semesters. A plug-in for Java editor recorded all changes made to the six course projects. Due to 1) the sheer size of the recorded data (up to 27k changes) and 2) that the data is temporal and multidimensional in nature, we've striven to summarize, aggregate, and visualize the data in ways easier to absorb.

Difference Estimation

In this study, we treated the students' Java programs as plaintext files to simplify analysis. We use the Levenshtein distance, which is the minimum number of edits required to change one text file into the other, as a metric to measure change over time and the change difference between students.

	P1	P2	P3	P4	P5	P6
Fall 2015	12	13	12	10	11	12
Fall 2016	7	5	10	8	10	10
Total	19	20	22	18	21	23
Changes	7k	6k	12k	7k	9k	3k

Table 1. Number of subjects participating in study for each year and each project; last row is approximate average number of changes for all subjects. Some of the recordings were corrupted, empty, or missing, and not all students in the class opted to participate.

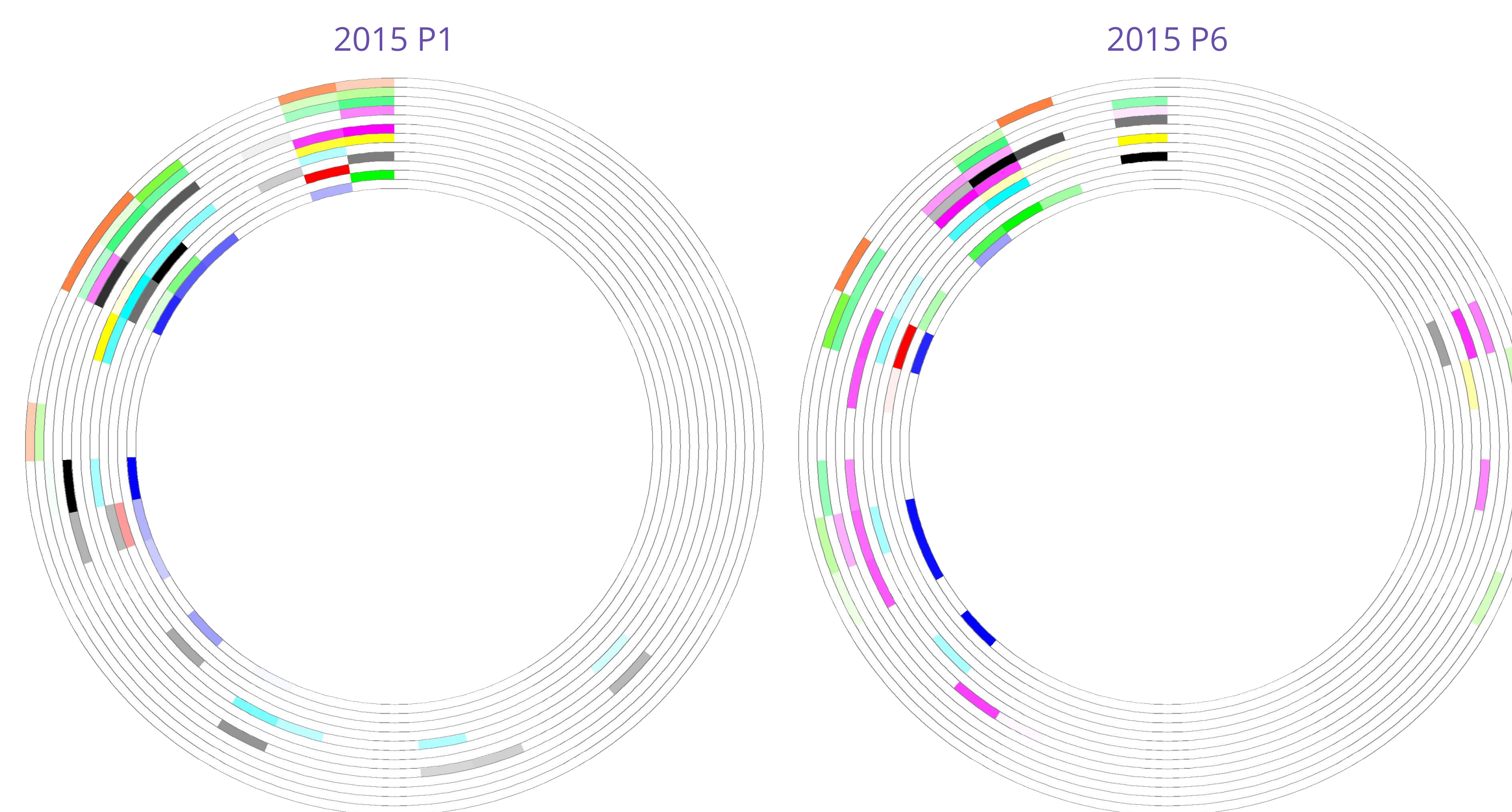


Figure 1. Donut chart showing when subjects worked on their project. The left subfigure is the first project of the course, and right subfigure is the last project. As seen in the left chart, many of the students did not start working on the assignment until close to the deadline, and they worked up to the end. The right chart shows several students started earlier, and many finished working before the deadline.

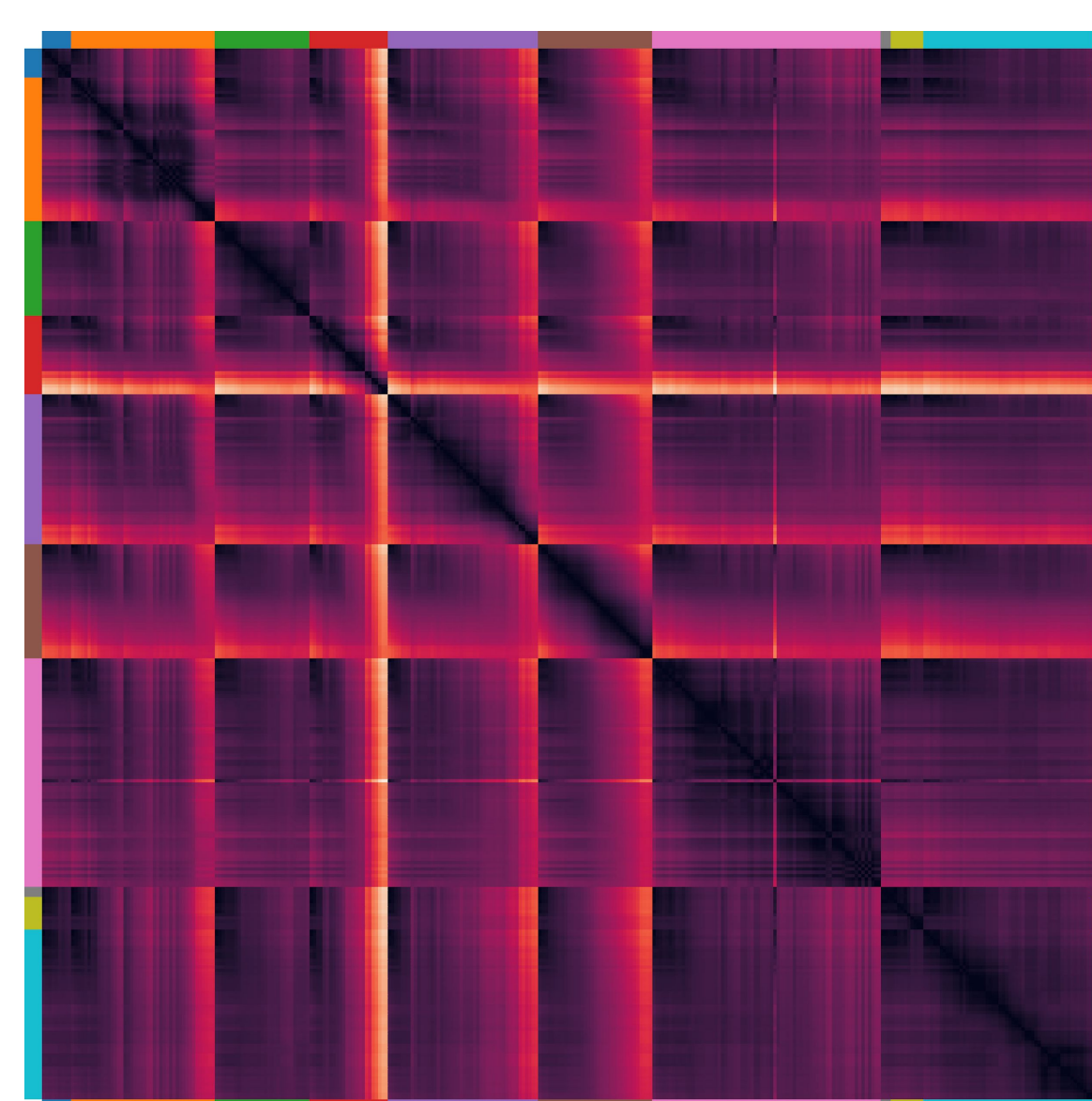


Figure 3. Heatmap of Project P5 for 2016 subjects. Comprehensive model; shows how different one snapshot of a student's work is compared to all other snapshots from all students throughout the project's lifespan (322 snapshots). Colored margins indicate rows and columns corresponding to each subject. Notice that block where cyan row and column intersect (bottom-right) is big and fairly cool, indicating the subject made very little difference in their source code over many changes.

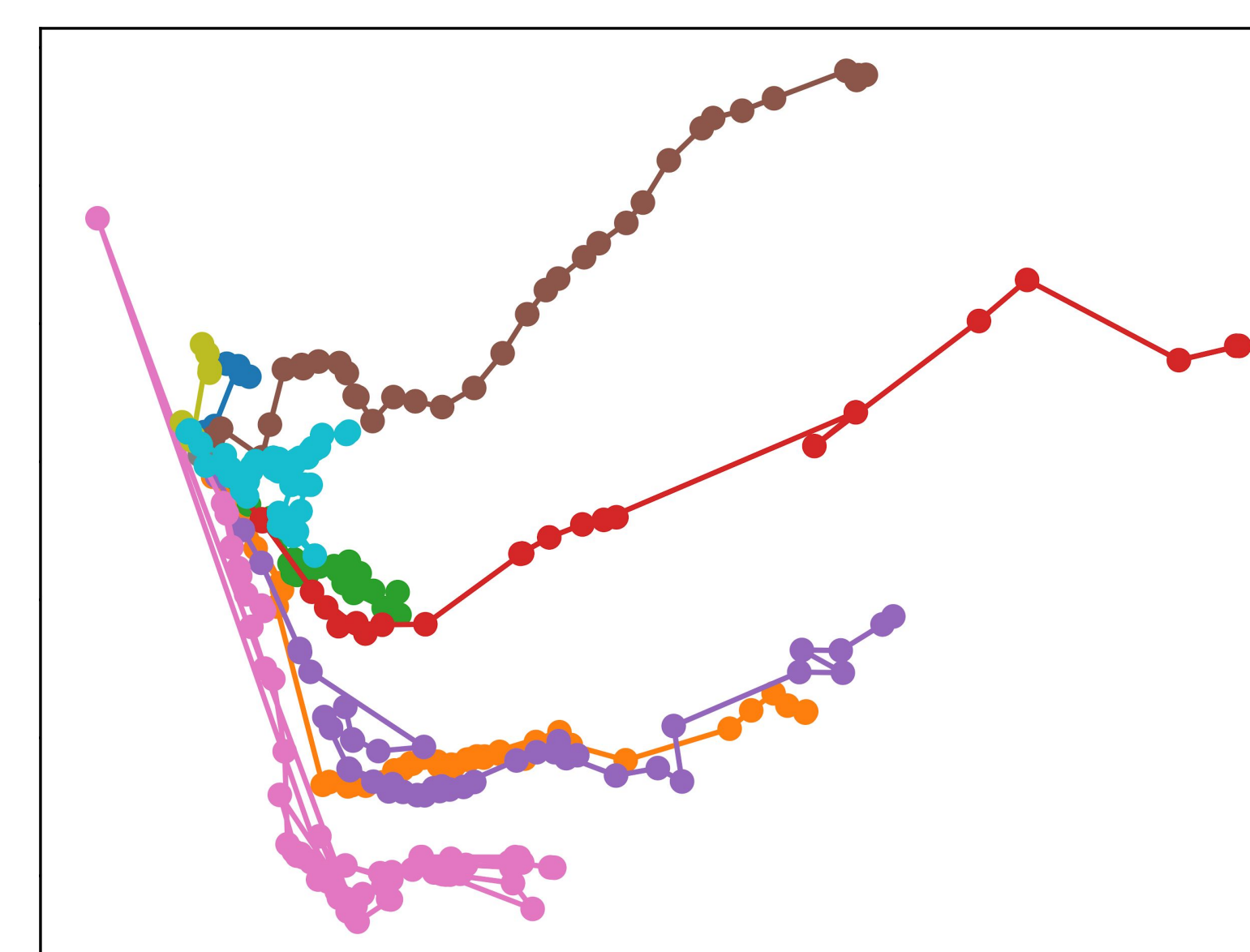


Figure 4. Dimensionally-Reduced Graph (DRG) of Project P5 for 2016 subjects. Lines represent subject's workflow, and line colors correspond to the colored margins of Figure 3. Nearby points indicate similar code. Parallel lines possibly indicate similar changes. Notice that the cyan line does not make much progress, indicating small difference made over many changes, while the red line ventures far, indicating much difference made in a few changes.

Figures

The donut charts (Fig. 1) illustrate the changes made over time, starting at the top and turning clockwise. Each ring represents a student. The highlighted areas represent differences, where large differences appear darker. The heatmap (Fig. 3) show the pairwise differences for snapshots, where cool and hot colors indicate small and large differences respectively. The DRG (Fig. 4) re-visualizes the rows from Fig. 3 into a line graph by reducing the dimensionality of the rows to two dimensions. Nearby points mean small differences, lines close together indicate similar changes, and divergent lines indicate different changes.

Conclusion

This semester project began the visualization explorations of the raw student data. Through these visualizations, we found a few interesting trends among the subjects (see figure captions). We believe our method can be applied to different domains to efficiently provide insights into subjects' workflows, and plan to explore other visualization techniques in the hopes of more formally publishing our findings in the future.

Acknowledgements

We would like to thank the subjects for their workflow data used in this research.