BMI 706 - Syllabus 2018

Course Director

- Nils Gehlenborg, PhD (nils@hms.harvard.edu)

Teaching Assistants

- Peter Kerpedjiev, PhD (pkerp@hms.harvard.edu)
- Sabrina Nusrat, PhD (sabrina_nusrat@hms.harvard.edu)

Time and Location

- Thursdays, 2:30 - 5:30 pm, Countway Room 403
- Seven classes from March 22nd through May 3rd

Piazza Class Management

- https://piazza.com/harvard/spring2018/bmi706

Learning Goals

1. Understand the role of visualization in the data analysis process.
2. Learn principles of perception and cognition in data visualization and become familiar with common visualization techniques and tools used in biomedical informatics.
3. Master critical evaluation of visualization approaches and become proficient in design processes for data visualization.
4. Acquire knowledge of programming techniques and tools and develop skills relevant for implementation of data visualizations.

Evaluation and Grades

Students will be evaluated on their class participation, quizzes, assignments, and final project, including presentation and report. Students are expected to complete required reading in time for the next session in order to be able to participate in the discussions.

Final Grade Composition
10% - Class Participation
10% - Weekly Quizzes (best 5 count)
30% - Assignments
50% - Project

Academic Integrity

1. Students must properly cite all submitted work appropriately.
2. Unless noted otherwise, students are expected to complete assignments, quizzes, and projects as individuals, not as teams. Discussion about course content and materials is acceptable, sharing of solutions is not acceptable.
3. You are encourage to consult websites for solutions to your coding problems but never copy code.
4. Disciplinary actions for "committing an act of academic dishonesty, such as plagiarism, cheating, or falsification of research results" (see Section 4.09 of the HMS Master's Student Handbook) are severe and typically will result in the requirement to withdraw from the program.

Attendance

Students must attend all classes unless they have explicit permission from the course director. Unexcused absence will affect the participation grade.

Quizzes

Quizzes will be online and distributed at the beginning of class. Please bring a computer to class!

Assignments

Assignments will be distributed via Piazza. Every student gets a total of 72 hours for late assignments.

Project

Students will have four weeks to work on their final project, beginning in Work 3. The homework assignments will help students to prepare for their project. Details about the project will be provided separately.

Reading

Required

- Visualization Analysis and Design, Tamara Munzner, CRC Press (2014)
- PDFs distributed by the instructor

Recommended

- Nature Methods Special Issue on Data Visualization (2010)
- Nature Methods Points of View columns (2010 - 2016)
- JAMIA Special Issue on Data Visualization (2014)
- IEEE InfoVis and IEEE VAST papers with bio/med/healthcare applications

Lectures

Session 1: Introduction to Data Visualization

- Course overview
- Role of data visualization
    - exploration, confirmation, communication
    - importance in the context of data analysis
    - data visualization as a research area
- Visual encoding: visual channels, color, pitfalls
- Overview of Plot.ly

Session 2: Design Process, Evaluation, and Interaction

Book Chapters

1. *Required:* Visualization Analysis & Design (VAD): Chapter 1, Chapter 4, Chapter 11
2. *Recommended* (Session 1 background): Chapter 2, Chapter 5

Papers (download from https://paperpile.com/shared/1q74pd)

1. *Required (*Presentation*):* Borkin, M.A., Gajos, K.Z., Peters, A., Mitsouras, D., Melchionna, S., Rybicki, F.J., Feldman, C.L., and Pfister, H. (2011). Evaluation of artery visualizations for heart disease diagnosis. IEEE Trans. Vis. Comput. Graph. 17, 2479–2488.
2. *Required:* Amar, R., Eagan, J., and Stasko, J. (2005). Low-level components of analytic activity in information visualization. In IEEE Symposium on Information Visualization, 2005. INFOVIS 2005., pp. 111–117.
3. *Recommended* (Session 1 background): Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods, William S. Cleveland and Robert McGill, Journal of the American Statistical Association, Vol. 79, No. 387 (Sep., 1984), pp. 531-554

- Design process
    - User-centered design, tasks, interviews, creativity workshops
- Evaluation
    - Qualitative, quantitative, challenges
- Interaction
    - lenses, zoom, animation

Session 3: High-Dimensional Data

Book Chapters

1. *Required:* Visualization Analysis & Design (VAD): Chapter 7

Papers (download from https://paperpile.com/shared/1q74pd)

1. *Required (Presentation):* Gratzl, S., Lex, A., Gehlenborg, N., Pfister, H., and Streit, M. (2013). LineUp: visual analysis of multi-attribute rankings. IEEE Trans. Vis. Comput. Graph. *19*, 2277–2286.
2. *Required:* Gehlenborg, N., O'Donoghue, S.I., Baliga, N.S., Goesmann, A., Hibbs, M.A., Kitano, H., Kohlbacher, O., Neuweger, H., Schneider, R., Tenenbaum, D., et al. (2010). Visualization of omics data for systems biology. Nat. Methods *7*, S56–S68.
3. *Required:* Gratzl, S., Gehlenborg, N., Lex, A., Pfister, H., and Streit, M. (2014). Domino: Extracting, Comparing, and Manipulating Subsets Across Multiple Tabular Datasets. IEEE Trans. Vis. Comput. Graph. *20*, 2023–2032.

- Multivariate data sets
  - homogeneous matrices: gene expression, protein abundance, etc.
  - heterogeneous matrices: multivariate rankings, etc.
- Interaction

Session 4: Networks

Book Chapters

1. Required: Visualization Analysis & Design (VAD): Chapter 9 + Chapter 12

Papers (download from https://paperpile.com/shared/1q74pd)

1. Required (Presentation): Barsky, A., Munzner, T., Gardy, J., and Kincaid, R. (2008). Cerebral: visualizing multiple experimental conditions on a graph with biological context. IEEE Trans. Vis. Comput. Graph. *14*, 1253–1260.
2. Optional: Barsky, A., Gardy, J.L., Hancock, R.E., and Munzner, T. (2007). Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. Bioinformatics *23*, 1040–1042.
3. Required: van den Elzen, S., and van Wijk, J.J. (2014). Multivariate Network Exploration and Presentation: From Detail to Overview via Selections and Aggregations. IEEE Trans. Vis. Comput. Graph. *20*, 2310–2319.
4. Required: Partl, C., Kalkofen, D., Lex, A., Kashofer, K., Streit, M., and Schmalstieg, D. (2012). enRoute: Dynamic path extraction from biological pathway maps for in-depth experimental data analysis. 2012 IEEE Symposium on Biological Data Visualization (BioVis) 107–114.

- protein networks, pathways, organizations, phylogenies

Session 5: Genomes and Epigenomes

Papers (download from https://paperpile.com/shared/1q74pd)

1. Required (Presentation): O'Brien, T.M., Ritz, A.M., Raphael, B.J., and Laidlaw, D.H. (2010). Gremlin: an interactive visualization model for analyzing genomic rearrangements. IEEE Trans. Vis. Comput. Graph. *16*, 918–926.
2. Required: Meyer, M., Munzner, T., and Pfister, H. (2009). MizBee: a multiscale synteny browser. IEEE Trans. Vis. Comput. Graph. *15*, 897–904.
3. Required: Nielsen, C.B., Cantor, M., Dubchak, I., Gordon, D., and Wang, T. (2010). Visualizing genomes: techniques and challenges. Nat. Methods *7*, S5–S15.
4. Required: Ferstay, J.A., Nielsen, C.B., and Munzner, T. (2013). Variant view: visualizing sequence variants in their gene context. IEEE Trans. Vis. Comput. Graph. *19*, 2546–2555.
5. Optional: Nielsen, C., and Wong, B. (2012a). Representing the genome. Nature Methods *9*, 423.
6. Optional: Nielsen, C., and Wong, B. (2012b). Managing deep data in genome browsers. Nat. Methods *9*, 521.
7. Optional: Nielsen, C., and Wong, B. (2012c). Representing genomic structural variation. Nat. Methods *9*, 631–631.

- genomes and epigenomes: browsers, SNPs, structural variants, copy number variation, 3D genome folding

Session 6: Time and Event Sequences

Papers (download from https://paperpile.com/shared/1q74pd)

1. Required (Presentation): Wang, T.D., Plaisant, C., Quinn, A.J., Stanchak, R., Murphy, S., and Shneiderman, B. (2008). Aligning Temporal Data by Sentinel Events: Discovering Patterns in Electronic Health Records. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (New York, NY, USA: ACM), pp. 457–466.
2. Required (please also consider for presentation): Wang, T.D., Plaisant, C., Shneiderman, B., Spring, N., Roseman, D., Marchand, G., Mukherjee, V., and Smith, M. (2009). Temporal summaries: supporting temporal categorical searching, aggregation and comparison. IEEE Trans. Vis. Comput. Graph. *15*, 1049–1056.
3. Required: Van Wijk, J.J., and Van Selow, E.R. (1999). Cluster and Calendar Based Visualization of Time Series Data. In Proceedings of the 1999 IEEE Symposium on Information Visualization, (Washington, DC, USA: IEEE Computer Society).
4. Required: Streit, M., and Gehlenborg, N. (2015). Points of view: Temporal data. Nat. Methods *12*, 97–97.
5. Optional: Monroe, M., Lan, R., Lee, H., Plaisant, C., and Shneiderman, B. (2013). Temporal event sequence simplification. IEEE Trans. Vis. Comput. Graph. *19*, 2227–2236.

6. Optional: Malik, S., Du, F., Monroe, M., Onukwugha, E., Plaisant, C., and Shneiderman, B. (2015). Cohort Comparison of Event Sequences with Balanced Integration of Visual Analytics and Statistics. In Proceedings of the 20th International Conference on Intelligent User Interfaces, (ACM), pp. 38–49.

- focus on healthcare data
  - EHR
  - Sensors such as activity trackers

Session 7: Final Presentations and Review

- Final presentations