# Internship: Grammar of Progressive Graphics

Advisor: Jean-Daniel Fekete, Email: Jean-Daniel.Fekete@inria.fr Web: http://www.aviz.fr/~fekete

Our civilization is collecting data at a pace never seen before. While data analysis has made tremendous progresses in scalability in the last decade, this progress has only benefited "confirmatory" analysis or model-based computation; progress in data exploration has lagged behind. The main reason is that, to maintain their efficiency during exploration, humans need a rapid feedback loop of about 10 seconds. However, when data becomes larger or algorithms more complex, bounding the latency while preserving accuracy is not possible with existing computation paradigms. To address this problem, a new computation paradigm is emerging: Progressive Data Analysis [Fekete&Primet2016], tying together the traditional imperative paradigm, stream processing, and sample-based approaches.

Data exploration, progressive or not, relies on data visualization to convey analysis results efficiently to humans, and to allow control over algorithms and visualizations through interaction. In the last years, several systems have been built according to the "Grammar of Graphics" described by Leland Wilkinson in his book [Wilkinson1999]: for the R system, GGPlot2 implements a "Layered Grammar of Graphics" [Wickham2010], and for Web-based visualization, Vega-Lite also implements a "high-level grammar of interactive graphics" [Satyanarayan et al.2017].

This internship will connect our ProgressiVis toolkit to Vega-Lite to study what are the extensions needed to the Grammar of Graphics to support progressive visualization and more generally progressive data analysis. Some extensions will be related to scalability: currently, Vega-Lite manages all the data in a web-browser that cannot handle large amounts of data. Data management and aggregation will need to be done at the level of the ProgressiVis toolkit before it is sent to Vega-Lite. Since the Vega-Lite specification is declarative, we expect that this separation of processing between the toolkit and the visualization can be done by modifying the implementation of the Vega-Lite language so as to keep it declarative. In addition to changes in the separation of processing, we know that some additions are needed to handle the progressiveness of the visualizations [Badam et al.2017]; exploring and specifying these additions relying on Vega-Lite and extensions will be the main goal of the internship.

**Location of the internship**: INRIA/Aviz, Bât 660, Paris-Sud

**As an Intern, you will be expected to:**

- collaborate with computer scientists from France, Korea, and the USA to integrate Vega-Lite with ProgressiVis

- create a functioning prototype of a progressive visualization system

- conduct scientific research, including literature studies, write a Master's thesis, and write a scientific article on the project.

**Requirements/skills:**

- you are a highly motivated student who is pursuing MS degree in data science, HCI, or related computer science topics,

- you have experience with software development, in Python and JavaScript,

- you have experience in modern visualization programming,

- you are able to clearly and concisely communicate in English in written and spoken form.

# References

1. [Fekete&Primet2016] Jean-Daniel Fekete, Romain Primet. Progressive Analytics: A Computation Paradigm for Exploratory Data Analysis. 2016.  ArXiv e-prints July 2016. URL: "http://arxiv.org/abs/1607.05162". 3, 10

2. [Wilkinson1999] Leland Wilkinson, The Grammar of Graphics, Springer, 2005.

3. [Wickham2010] Hadley Wickham, "A layered grammar of graphics", Journal of Computational and Graphical Statistics, vol. 19, no. 1, pp. 3-28, 2010.

4. [Satyanarayan et al.2017] Arvind Satyanarayan, Dominik Moritz, Kanit Wongsuphasawat, Jeffrey Heer, Vega-Lite: A Grammar of Interactive Graphics, IEEE Trans. Visualization & Comp. Graphics (Proc. InfoVis), 2017

5. [Badam et al.2017] Sriram Karthik Badam, Niklas Elmqvist, Jean-Daniel Fekete, Steering the Craft: UI Elements and Visualizations for Supporting Progressive Visual Analytics, Computer Graphics Forum, Wiley, 2017, Eurographics Conference on Visualization (EuroVis 2017), 36 (3), pp.12.