With a growing amount of available data, the approach we take in working with and investigating this data is of paramount importance. While the scientific method is underlying for data science as well, a modern approach to solving data-based problems should be more iterative since the issue of having too much data is becoming as common as having too little data. In this work we describe an agile approach to data science called lean data science and give examples of how to approach problems this way. We then describe our work on two specific problems, namely inferring other data sources from WiFi data and effectively scaffolding educational peer review. From The Copenhagen Network Study [Stopczynski et al., 2014] we have been able to work with a dataset collected using more than 1,000 smartphones from students over multiple years. Using this dataset we look at how well WiFi scans are able to replace other data sources such as Bluetooth and GPS. We show that WiFi data can accurately detect so-called stop-locations of the same quality as state-of-the-art GPS-based methods and that WiFi data can mirror Bluetooth data for the purpose of detecting face-to-face interactions between people. Peergrade is a web-based system for facilitating educational peer review built by us. Conceptually the idea of having students review work by other students serves many purposes, including a potential for saving time on grading for teachers and a way to train taxonomically complex skills for students. Because peer review is a complex process (many people reviewing many people) and requires things such as anonymity, evaluation criteria and the ability for instructors to moderate, it is best facilitated using a digital tool. We first describe Peergrade and some of the features it offers to educators. We then describe different research projects around Peergrade that attempts to help quantify the quality of reviews, allocate reviewers in an optimal way and automatically flag problematic feedback for moderation.