Introduction

Knock Software's *Ride Report* app combines a simple thumbs-up/thumbs-down rating system with GPS traces of bicycle rides to compile a crowdsourced data set of which routes are and are not stressful for urban bicyclists.

The app that collects the data is simple: *Ride Report* automatically detects when a user start riding their bike, records the GPS trace of the route, and then prompts the user at the end of the ride to give either a thumbs-up or thumbs-down rating. From this, they were able to create a simple "stress map" of Portland, OR, which displays the average ride rating of rides going through each discretized ride segment.

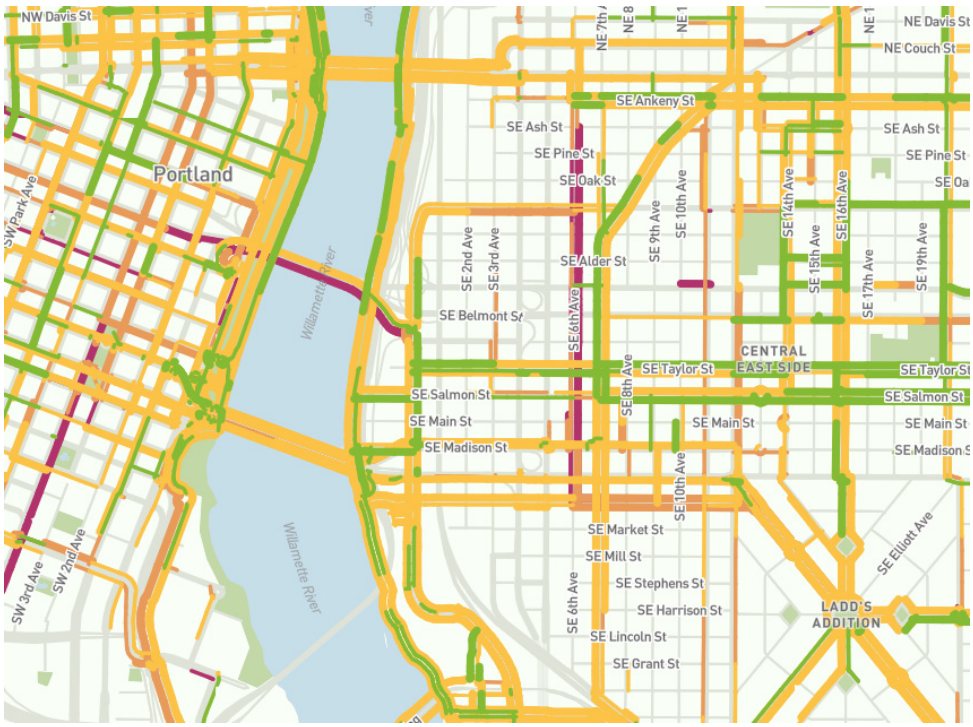

Figure 1: Ride Report's Stress Map for Portland. Greener road segments incidate less while more red segments indicate more stressful streets. "Stress" computed by taking the average rating for each segment.

The app privileges reducing barriers to response to increase sample size over ensuring quality and consistent responses. This presents the first problem: how can we analyze ratings when riders are likely rating rides inconsistently?

At the same time, we have another challenge. We have ratings associated with routes, but we would like to know the effect of particular road segments, for both inference (what effect does this road segment have on the rating?) and prediction (given a route, what do we expect the rating to be?) purposes.

Finally, there are interesting issues with missing data. First, the sample of rides we have are biased towards routes that riders percieve as better. Second, a significant portion of the rides are missing a response, and non response is unlikely to be independent of the response. The good news is that we have all the predictors for every ride, enabling us to build a model that is able to leverage the data with missing responses rather than listing the missing data as a liability.

Accounting for Rider Rating Variance

For ratings we are interested in modeling variance between riders (as we might expect riders to have different rating patterns than their peers) and within riders (as riders may not rate the same route and conditions the

same every time). To model this, we propose using multilevel regression, with random effects from each rider. This approach has been used in similar situations, in one case to model sexual attraction[1].

In a multilevel model, we fit a regression where the intercept term varies by group but comes from a common distribution. For example, if we let r_i be the rating of the ith ride, X_i be the ride-level variables, then we can fit a regression:

$$\mathbb{P}(r_i = 1) = \text{logit}^{-1}\left(\alpha_{j[i]} + X_i\beta\right),$$

where α_j is an intercept specific to rider j. In addition, the rider intercepts come from a common distribution,

$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2),$$

where μ_α is the mean of all the α_js.

Using varying intercepts will allow our model to exhibit the property of varying rates of riders giving stressful ratings. This can be extended to model differences in how riders respond to different kinds of road conditions using varying slope models. We explore multilevel model further in Section **??** and multilevel models for riders in Section **??**.

Addressing Road Segments as a Level

We examine multiple approaches to modeling road segments. In the first, we regard add regression term for the route, that is a weighted sum of terms for each road segment in the route, weighted by the lengths of the segments. The terms for the

the contribution of the route to be a weighted sum of the contributions of road segments in the route, weighted by the lengths of the segments.

Approaching Missing Data at Multiple Levels

This data set contains missing data at two levels.

First, there are many routes without any ratings. The routes taken by riders are not a random sample of routes, but are often already chosen by the rider as the least stressful ride. As we might expect because of this bias, only a small proportion of rides are rated as "stressful."

One solution we explore to this problem is adding a segment popularity parameter as a segment-level predictor.

Second, not every ride is given a rating. We do have the route they chose, and all associated covariates, but the response variable, rating is missing. As we will discuss in chapter **??**, the pattern of non-response is unlikely to be unbiased.

To address the second problem, we first build a separate probability model that predicts non response, and then integrate that model into our main model.

[1] @mackaronis2013