Intro to Data Science – Miniproject

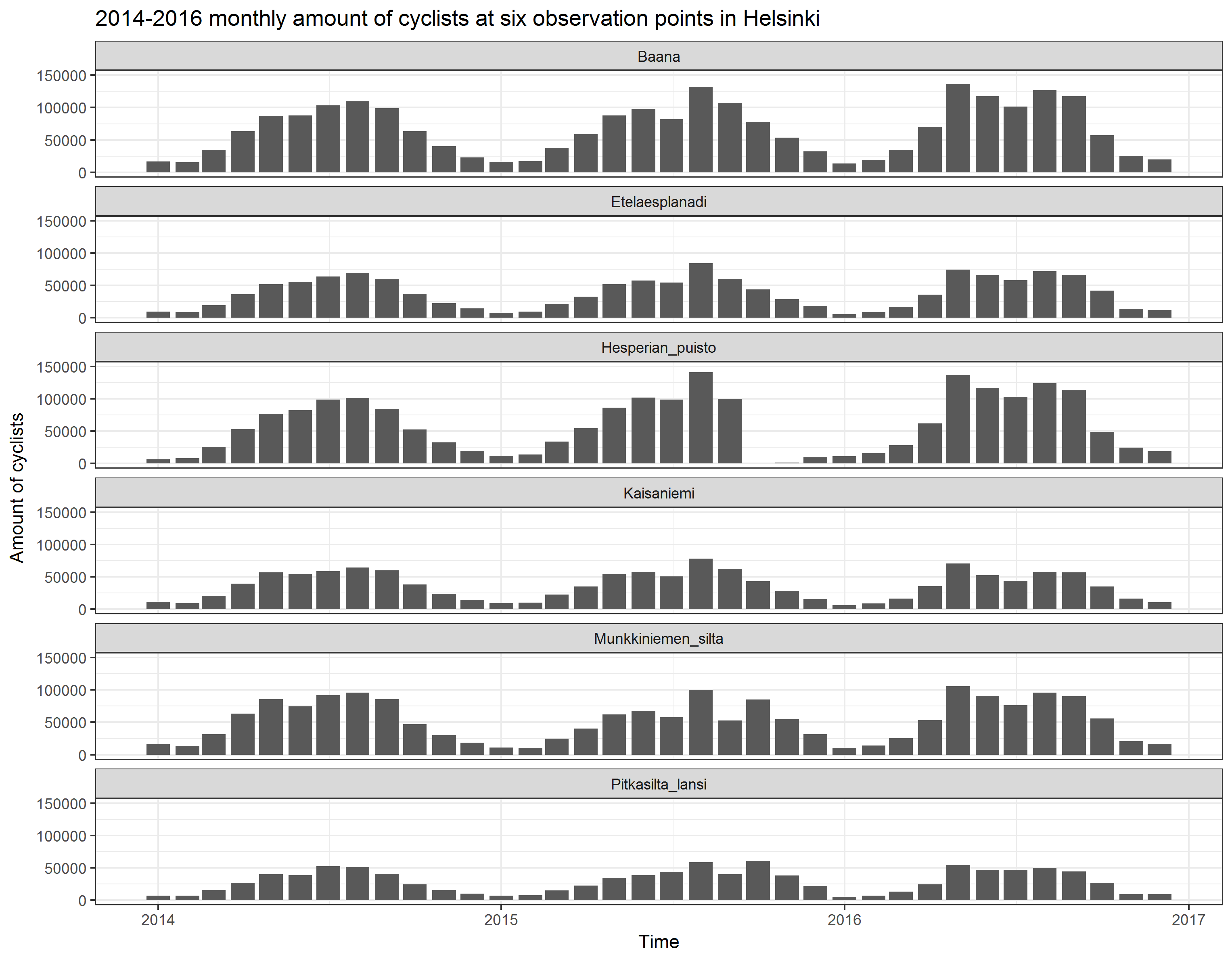
Mikko Laapas, Timo Mäki, Jani Rapo

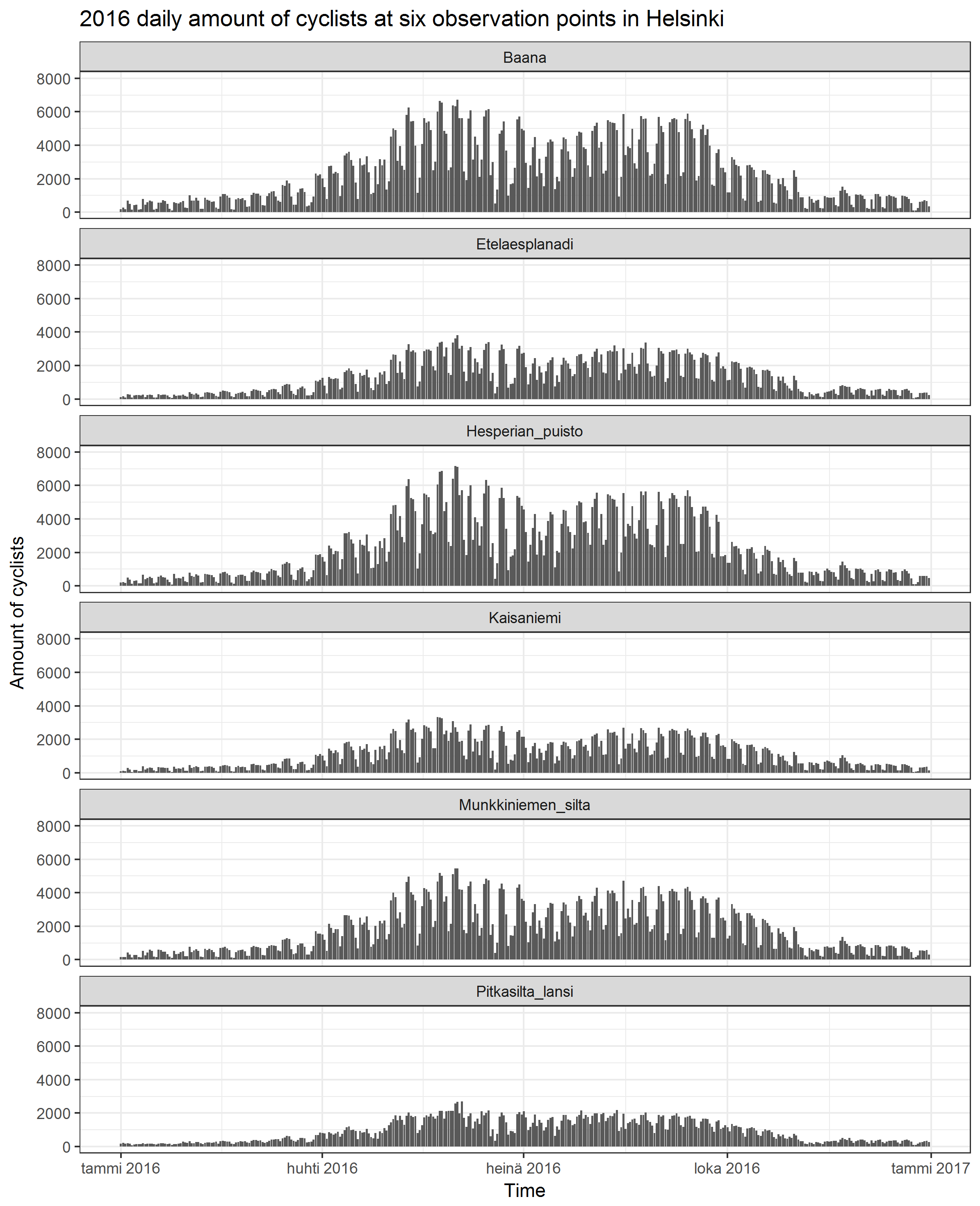
# Variations in bicycle data – does weather have a role?

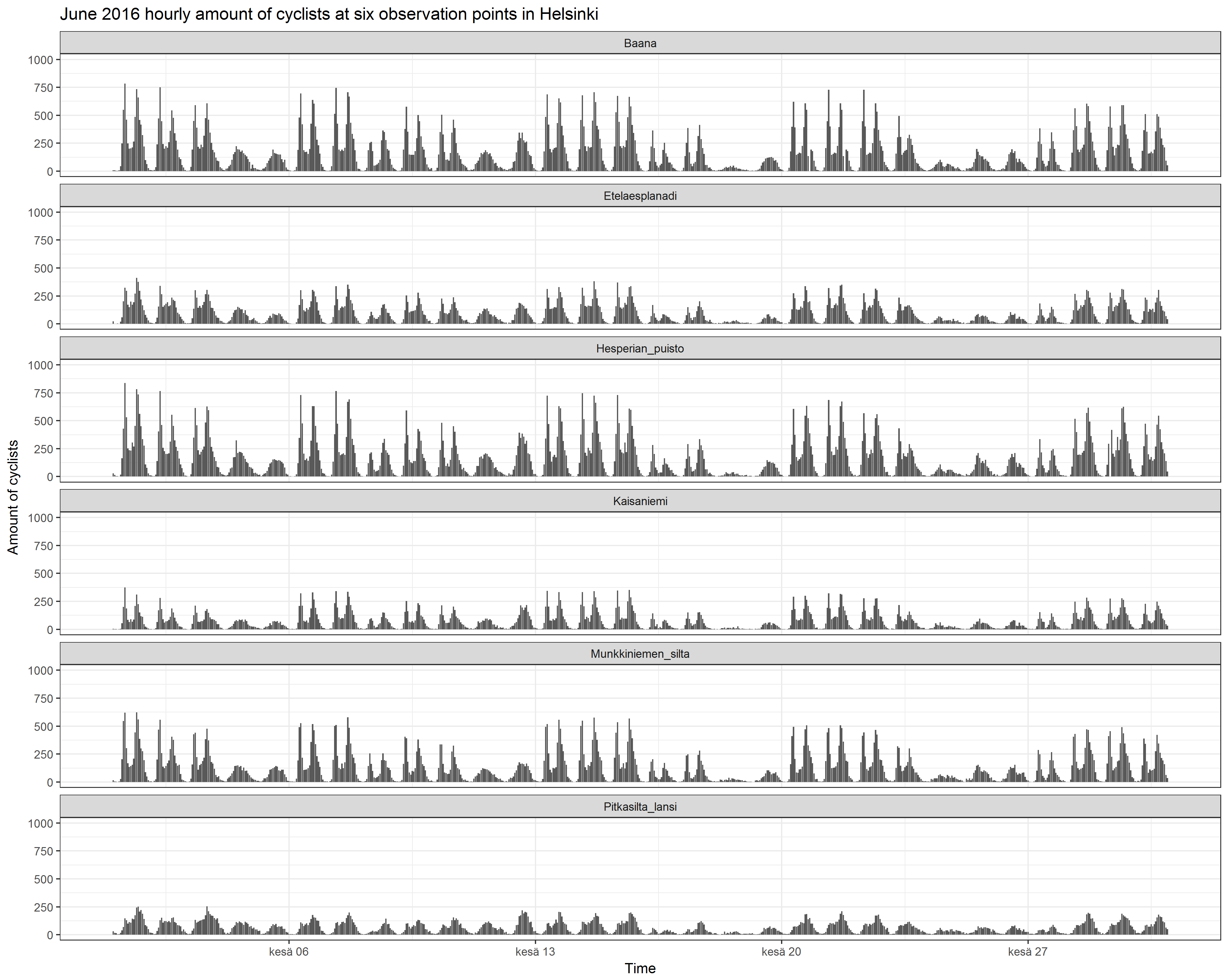
Besides weather conditions, there are three rather obvious factors causing variations in bicycle data:

* Time of year/seasonal/monthly variation based on the seasonality of Finnish climate
* Time of day variation based on the day-night cycle
* Weekday variation based on the more common working days Monday to Friday and weekends

These three variations are presented in following examples using 2014-2016 cyclists data from 6 observation points around Helsinki.



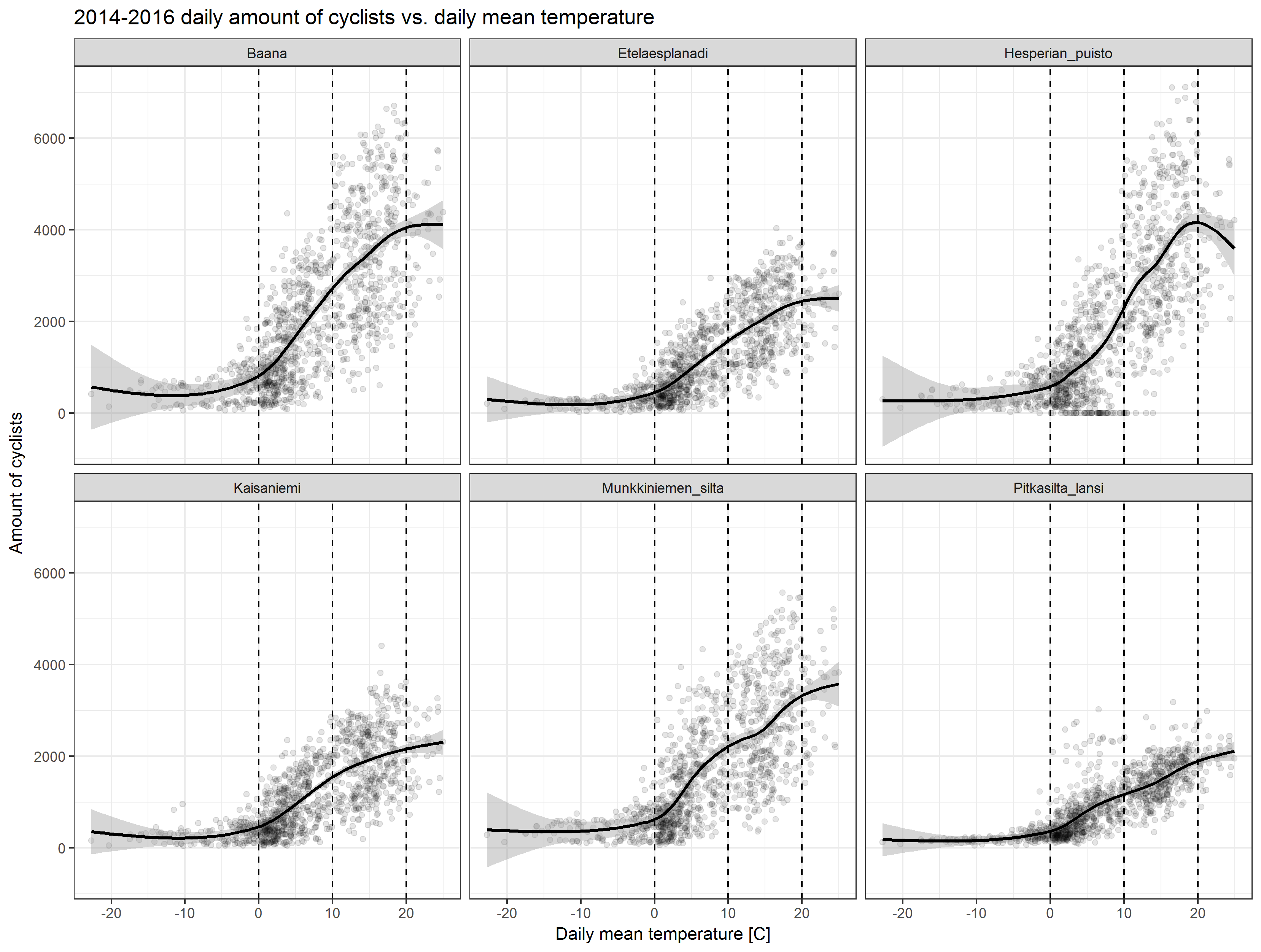




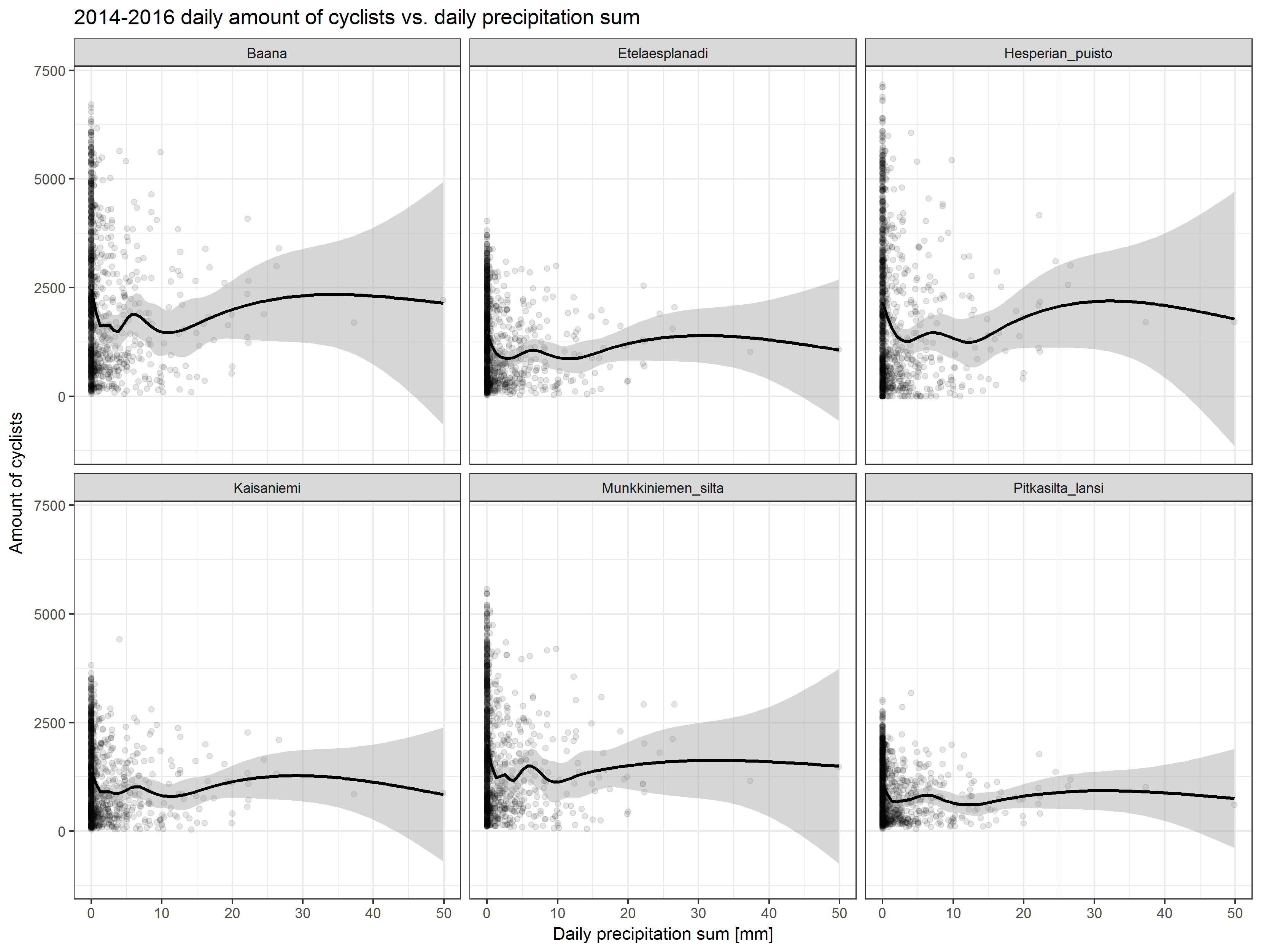
Some observations:

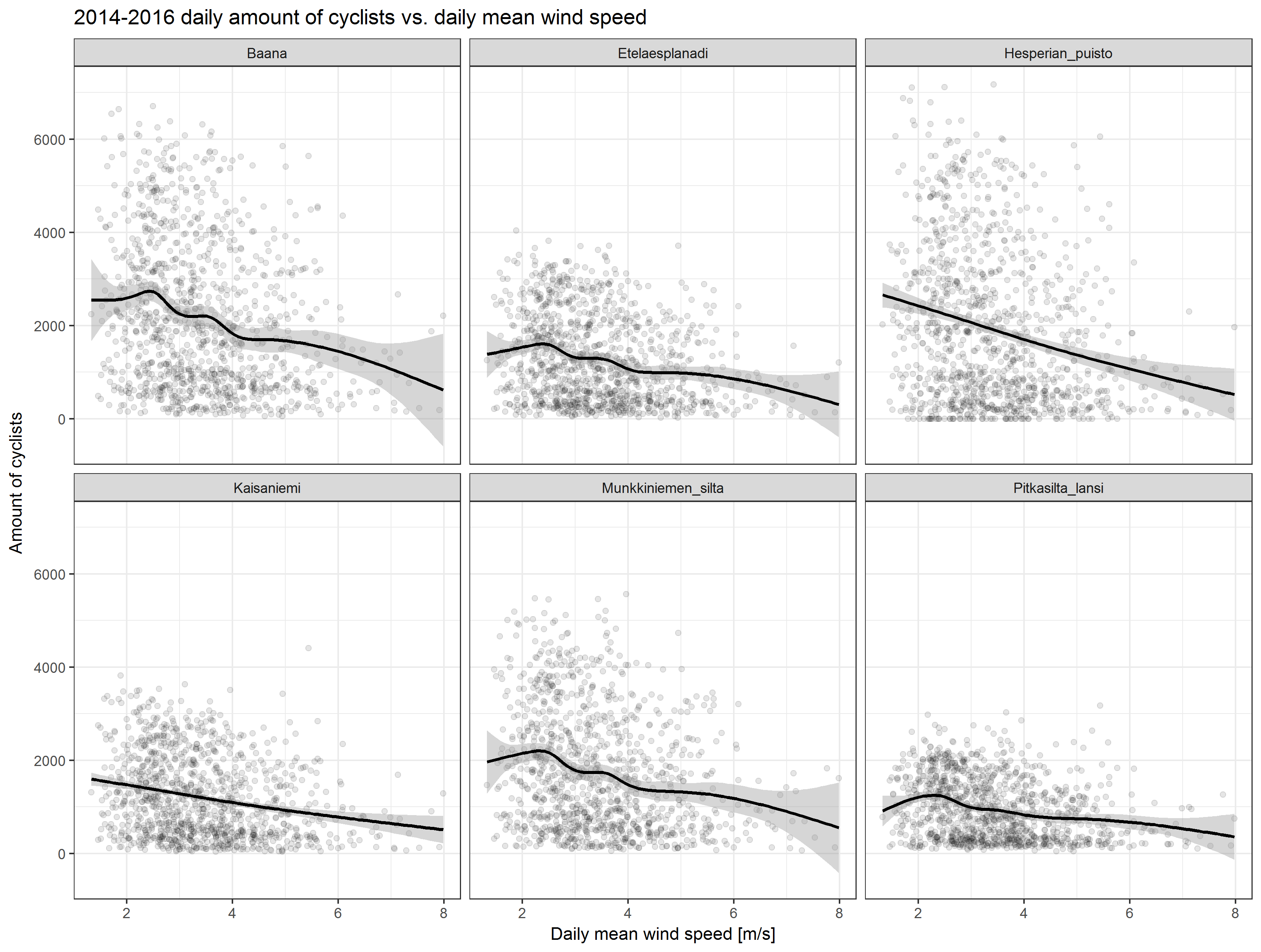
* At hourly level, there’s two maxima during work commute hours
* At daily level working week weekend cycle is clear. People commuting to work not necessarily cycle during weekends.
* July as the most common summer holiday month shows in the data.
* Indeed, Helsinki is empty during Midsummer festivities.

# Relationship between weather variables and amount of cyclists



* Drastic drop in cyclists when temperature drops below zero
* Thresholds at 0, 10 and 20 degrees?
* Below zero (or -5), the amount of cyclists remain rather steady. These are the hardcore cyclists who ride their bikes whatever the conditions.
* When daily mean temperature is over 20, it is starting to be too hot for cycling?



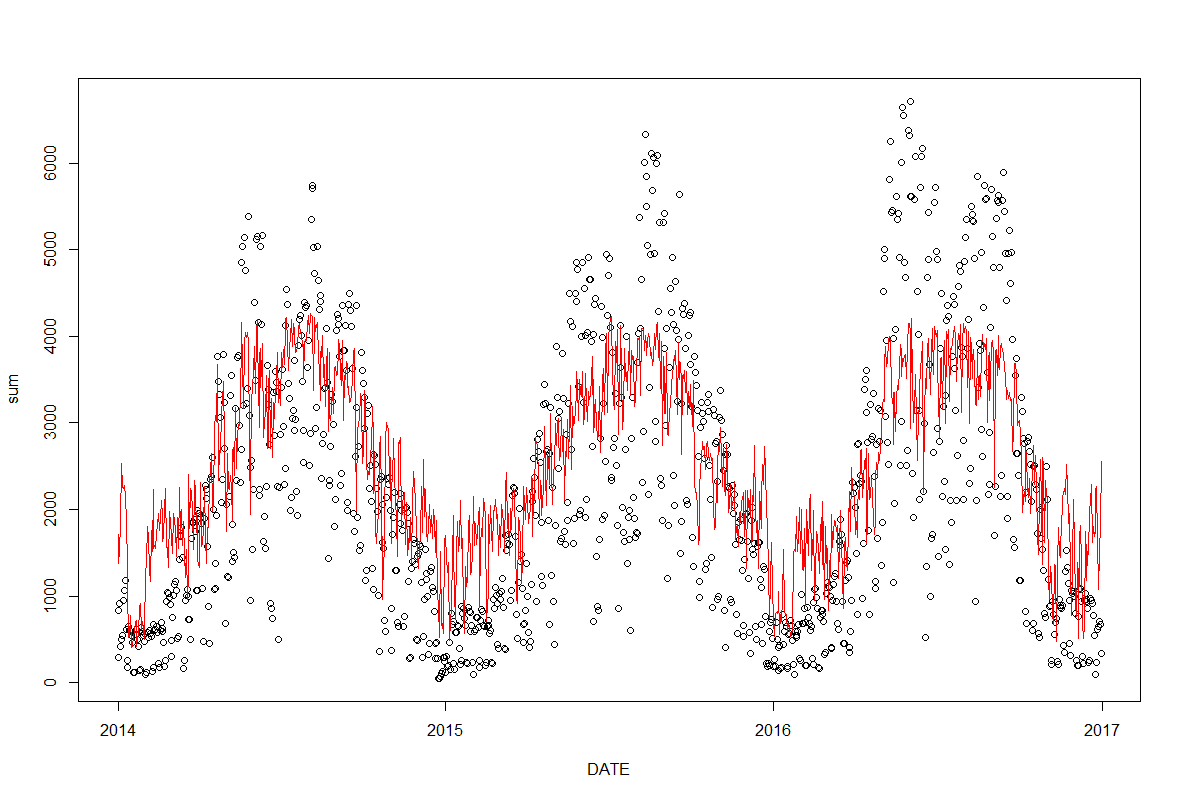


* Daily precipitation and windiness seems also to have effect:
  + Most cyclists during non-rainy days
  + Amount of cyclists decline during windy days

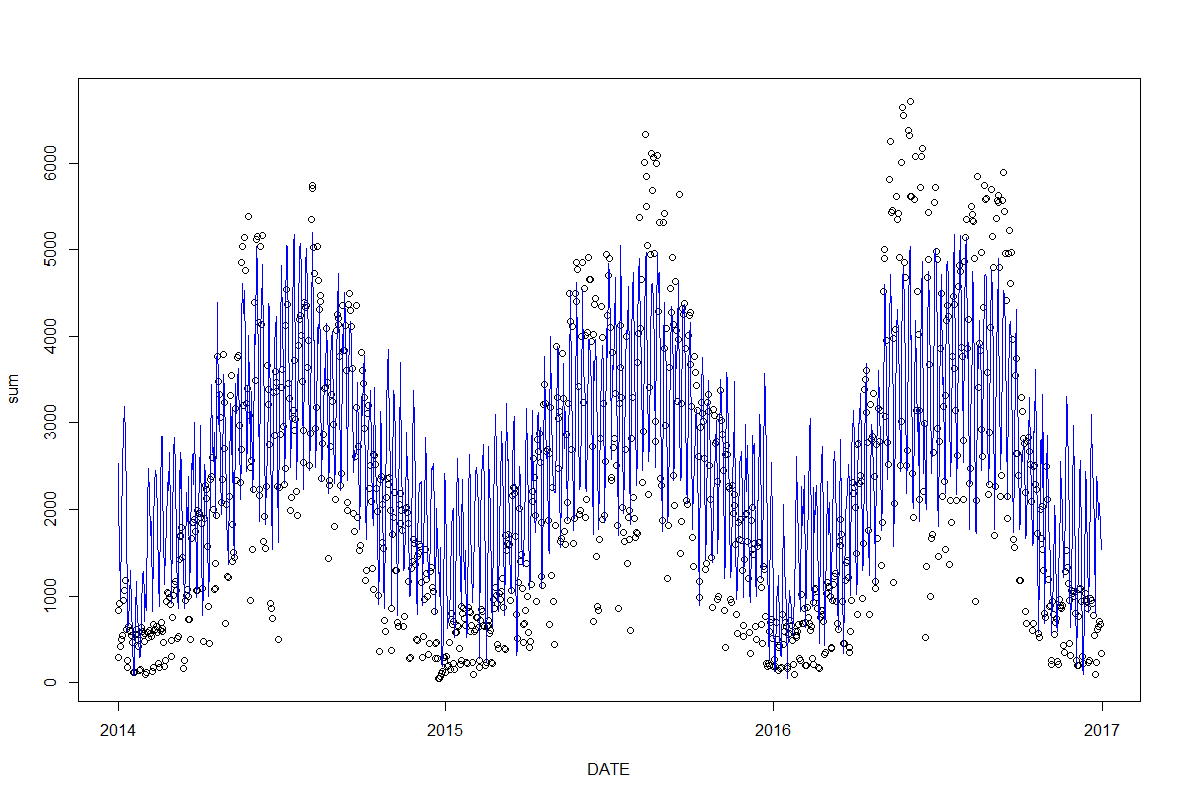
# Loess fitting with different predictors – case Baana

cyclists ~ temp + wind + prec

* variability and seasonal cycle too small



cyclists ~ temp + wind + prec + wday

* better, but still overestimates winter and underestimates summer
* 

R’s *loess* takes only max 4 predictors, so let's change prec (lowest correlated) to month of year

* Best fit so far (cor obs vs pred 0.89), but there’s still something going on during summer

