

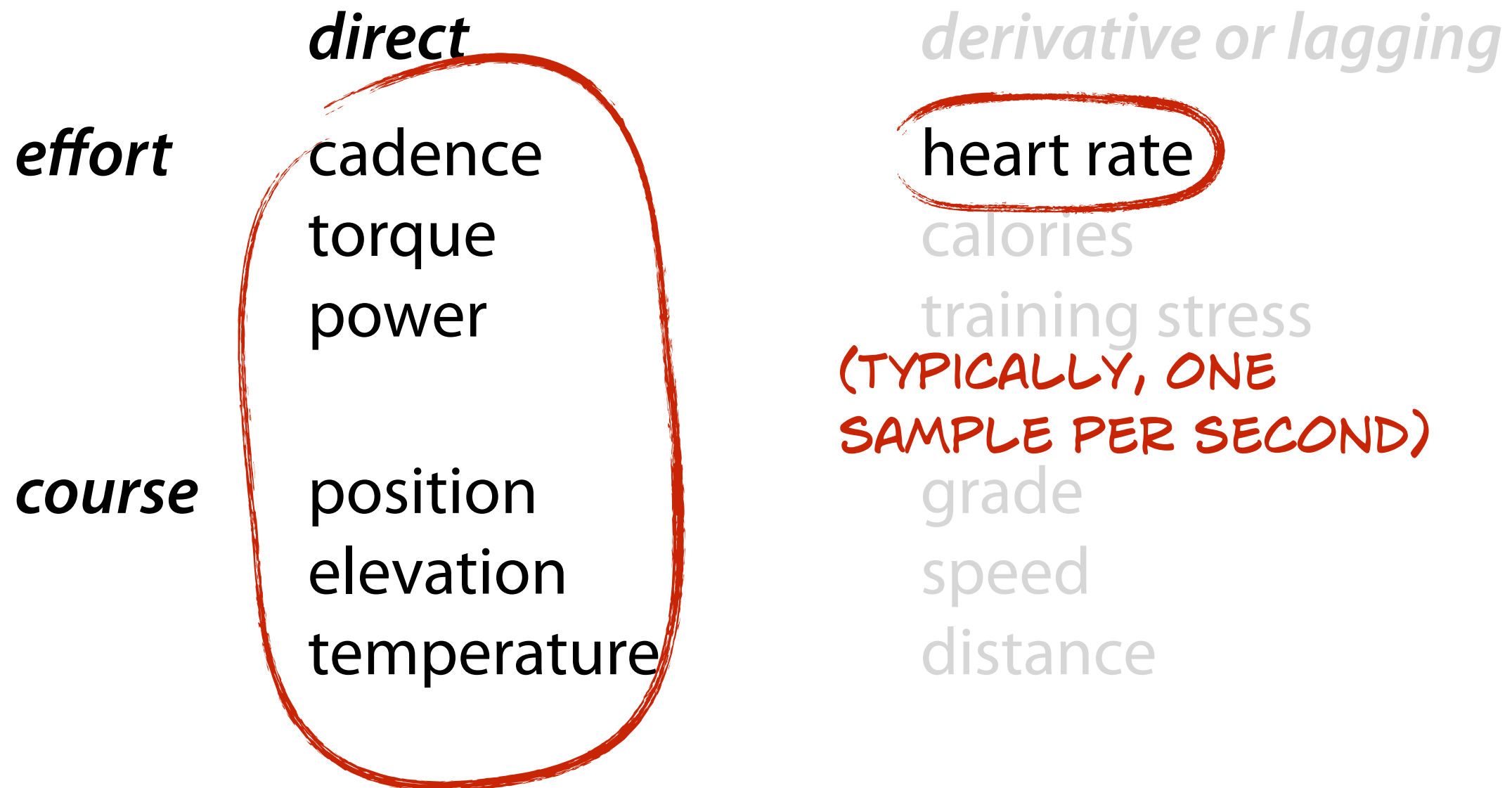
# Analyzing endurance sports activity data with Spark

**William Benton**  
**Red Hat, Inc.**  
**[willb@redhat.com](mailto:willb@redhat.com)**

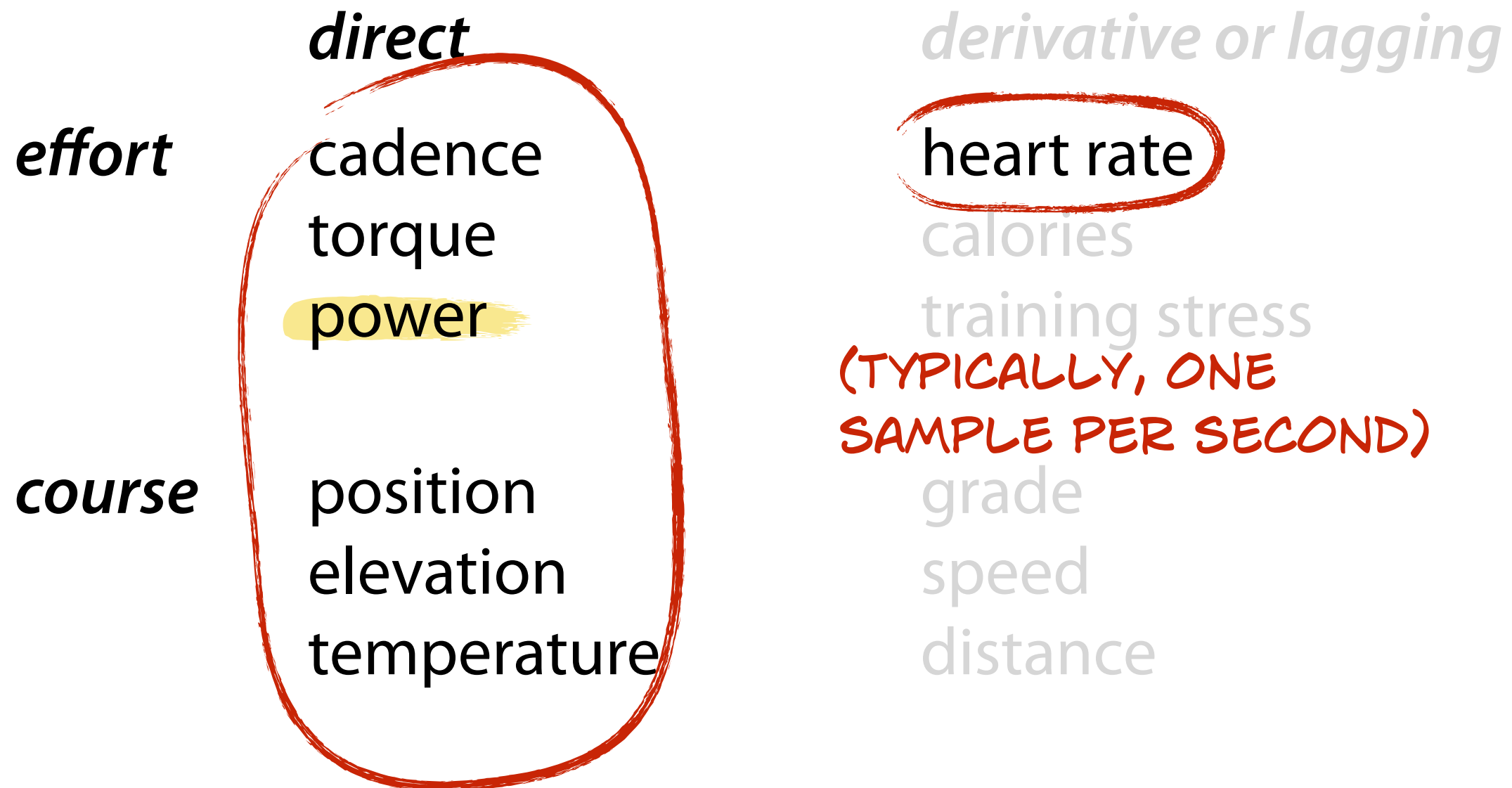
# Metrics available to cyclists

	<i>direct</i>	<i>derivative or lagging</i>
<i>effort</i>	cadence torque power	heart rate calories training stress
<i>course</i>	position elevation temperature	grade speed distance

# Metrics available to cyclists



# Metrics available to cyclists





**“Where should  
I do intervals?”**

**MT. DIABLO (NORTH GATE TO SUMMIT)**  
NEAR WALNUT CREEK, CA

**3,970'**  
**10.5 MILES**

**405'**

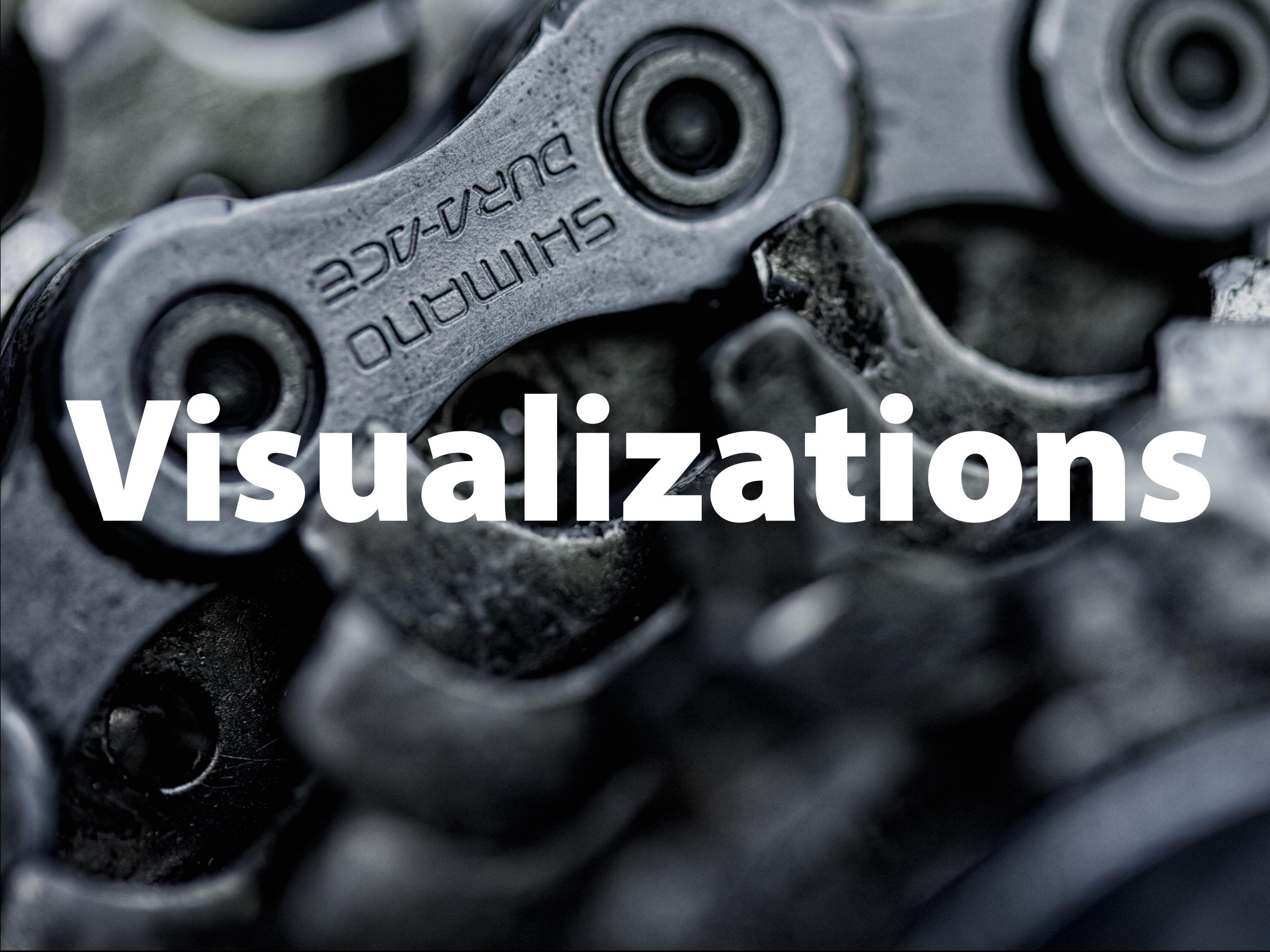
**"BUDDHA'S PALM"**  
NEAR CROSS PLAINS, WI

**1,214'**

**890'**

**16.8 MILES**



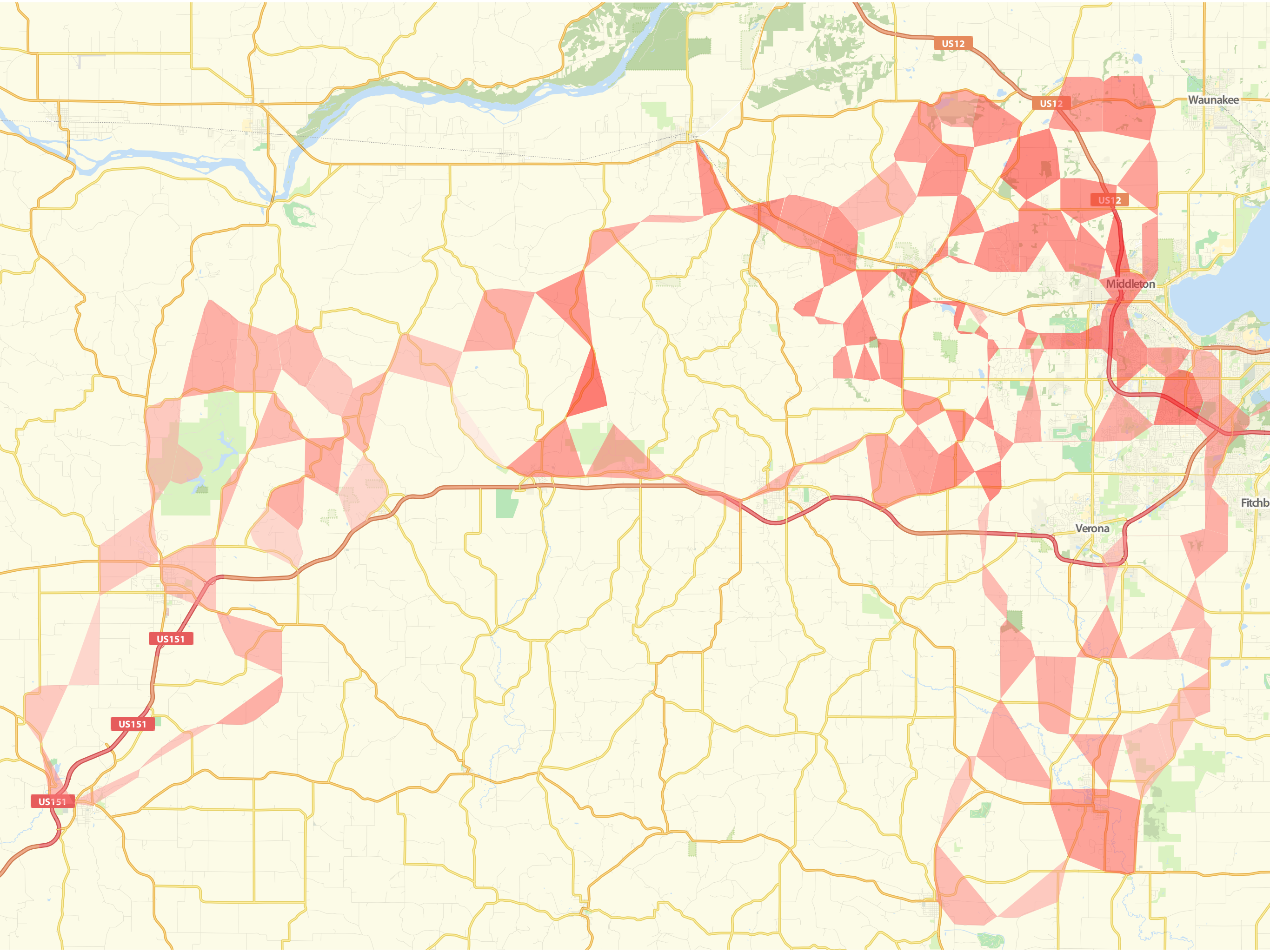


# Visualizations

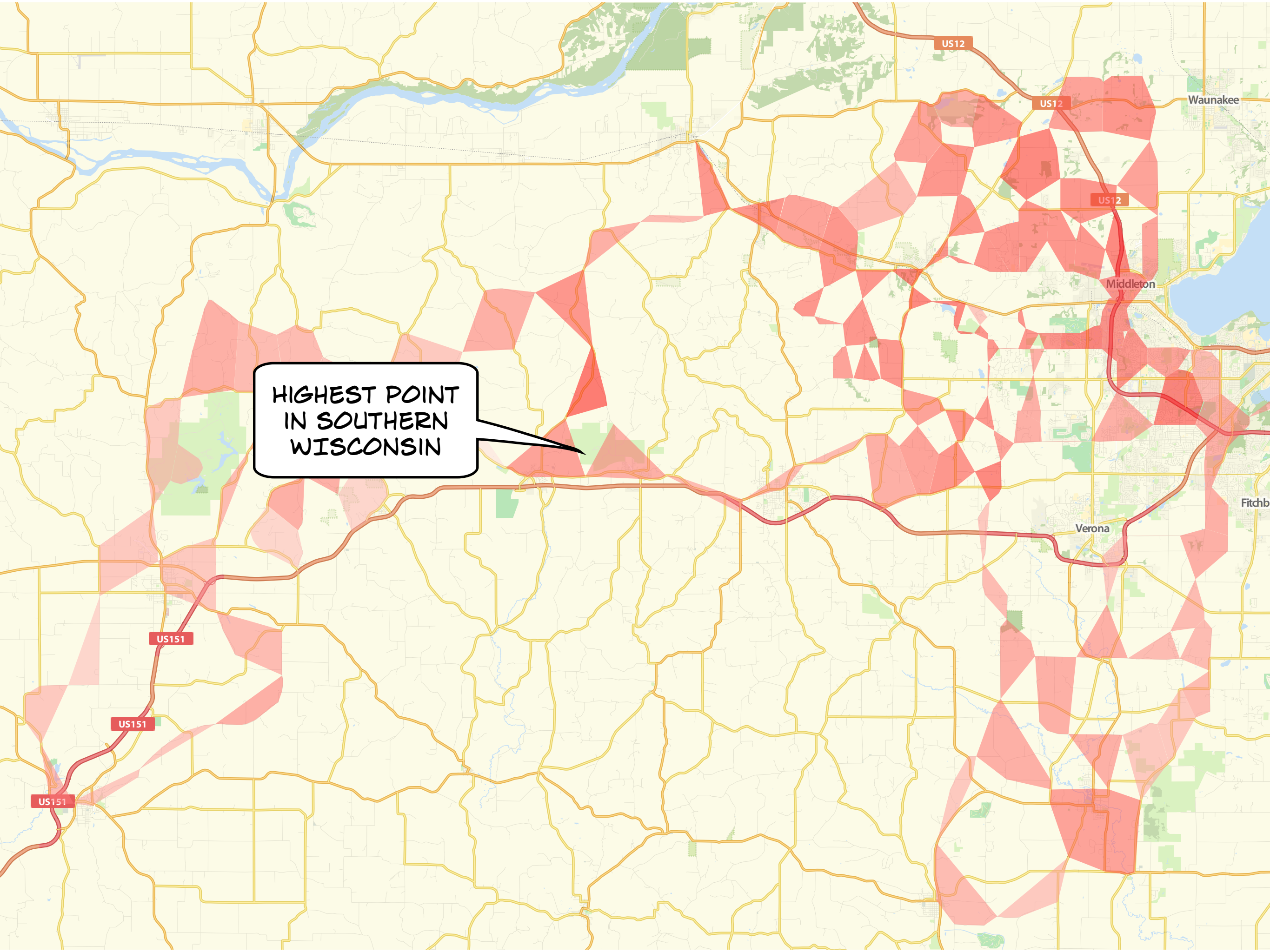


# **“Power heatmap”**

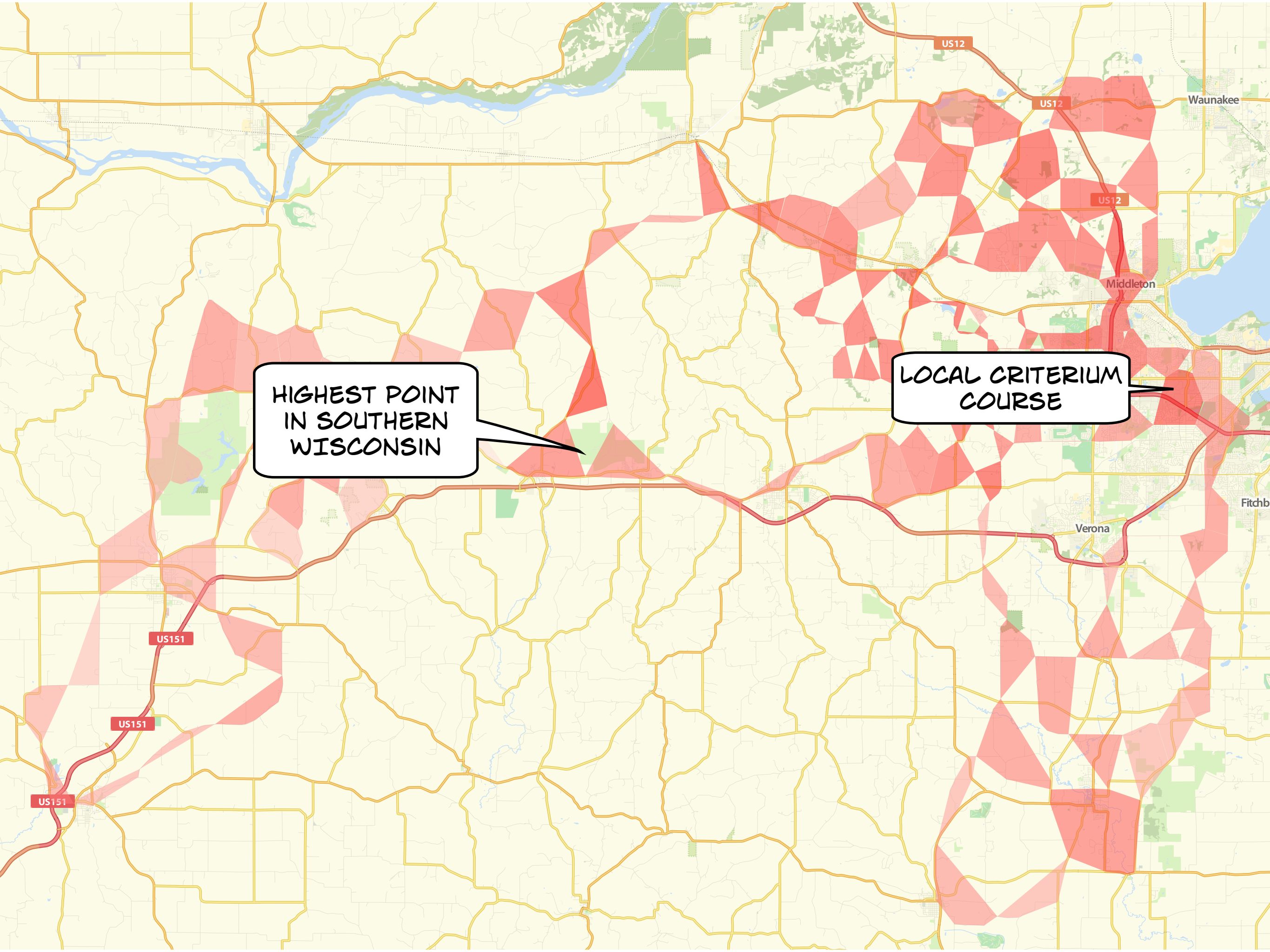
- **Cluster points from all activities**
- **Find the best ten-minute effort starting in each cluster**
- **Color the convex hull of each cluster based on best effort quality**







HIGHEST POINT  
IN SOUTHERN  
WISCONSIN

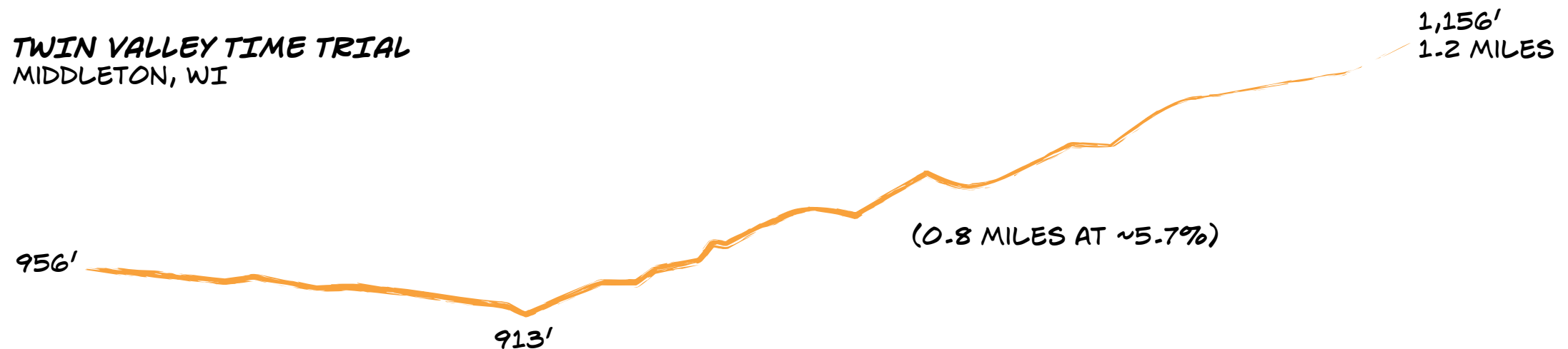


HIGHEST POINT  
IN SOUTHERN  
WISCONSIN

LOCAL CRITERIUM  
COURSE



# Finding best efforts



***A four-minute all-out effort includes sixty strong three-minute efforts.***



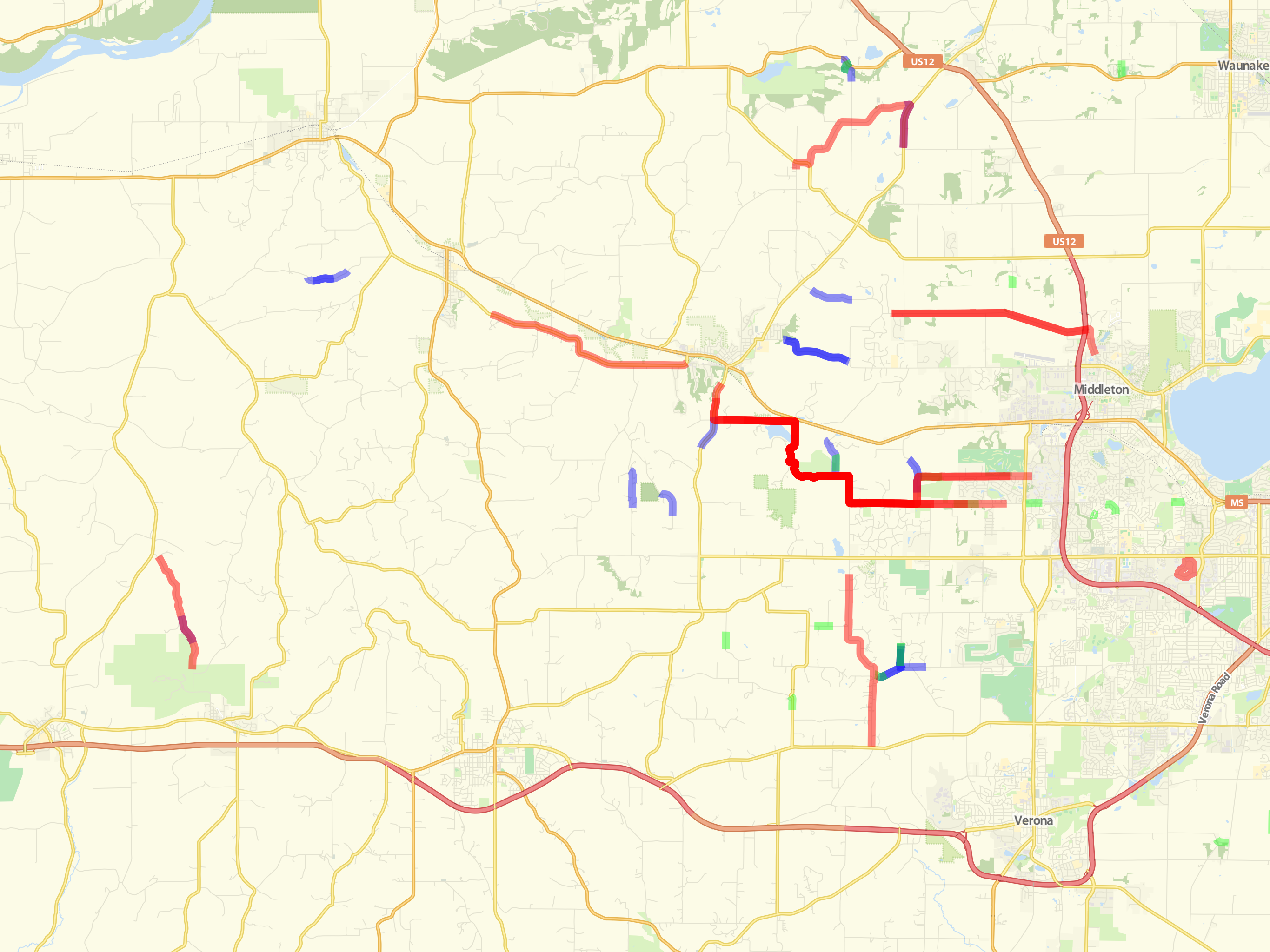
# Finding best efforts

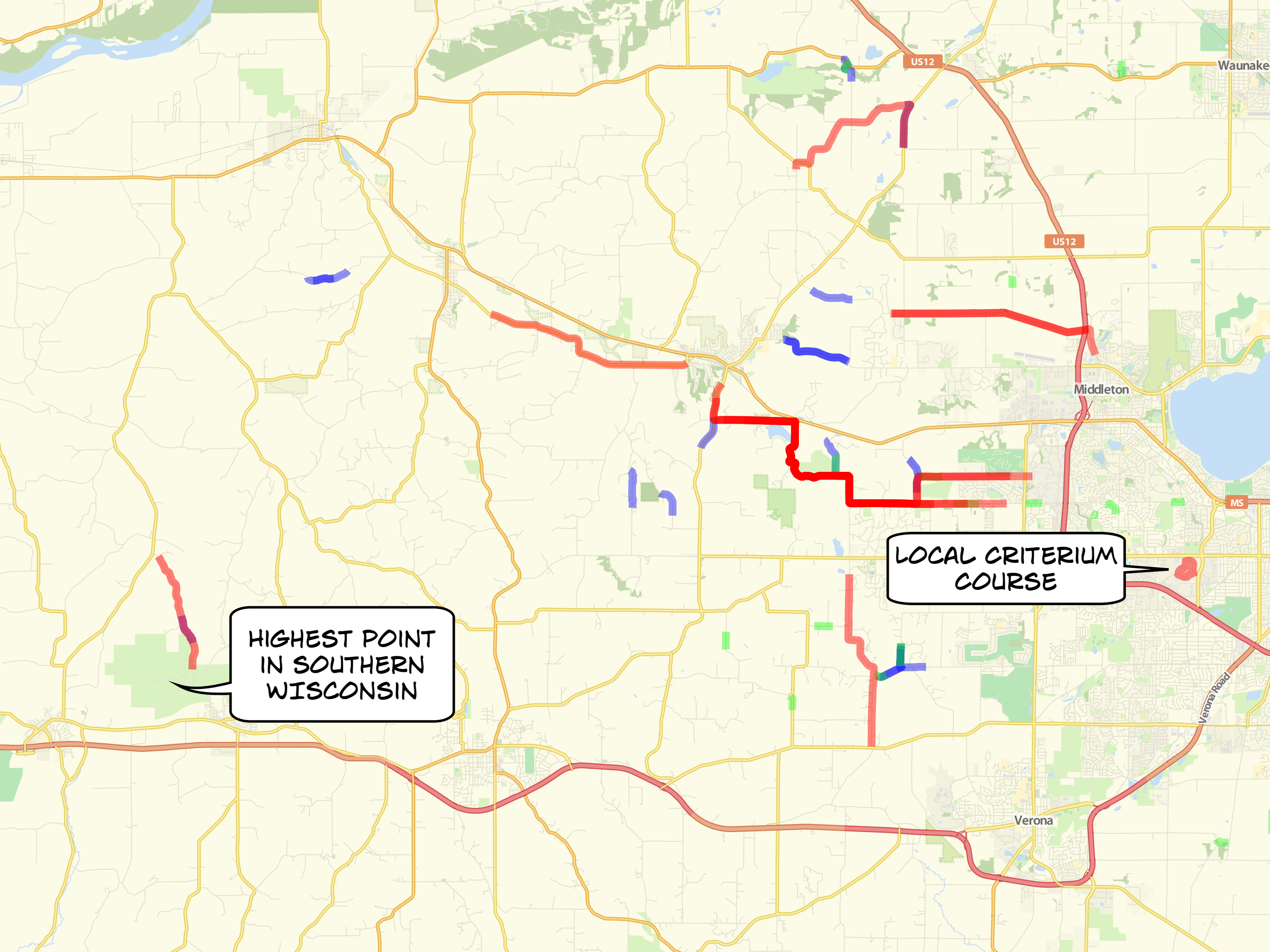
- Interval workouts can confound the naïve approach
- For each pair of spatial clusters  $(i, j)$ , find the best effort between  $i$  and  $j$
- Plot best efforts for each pair

# Finding best efforts

- Interval workouts can confound the naïve approach
- For each pair of spatial clusters  $(i, j)$ , find the best effort between  $i$  and  $j$
- Plot best efforts for each pair

(11 RDD TRANSFORMATIONS; 2 RDD ACTIONS)

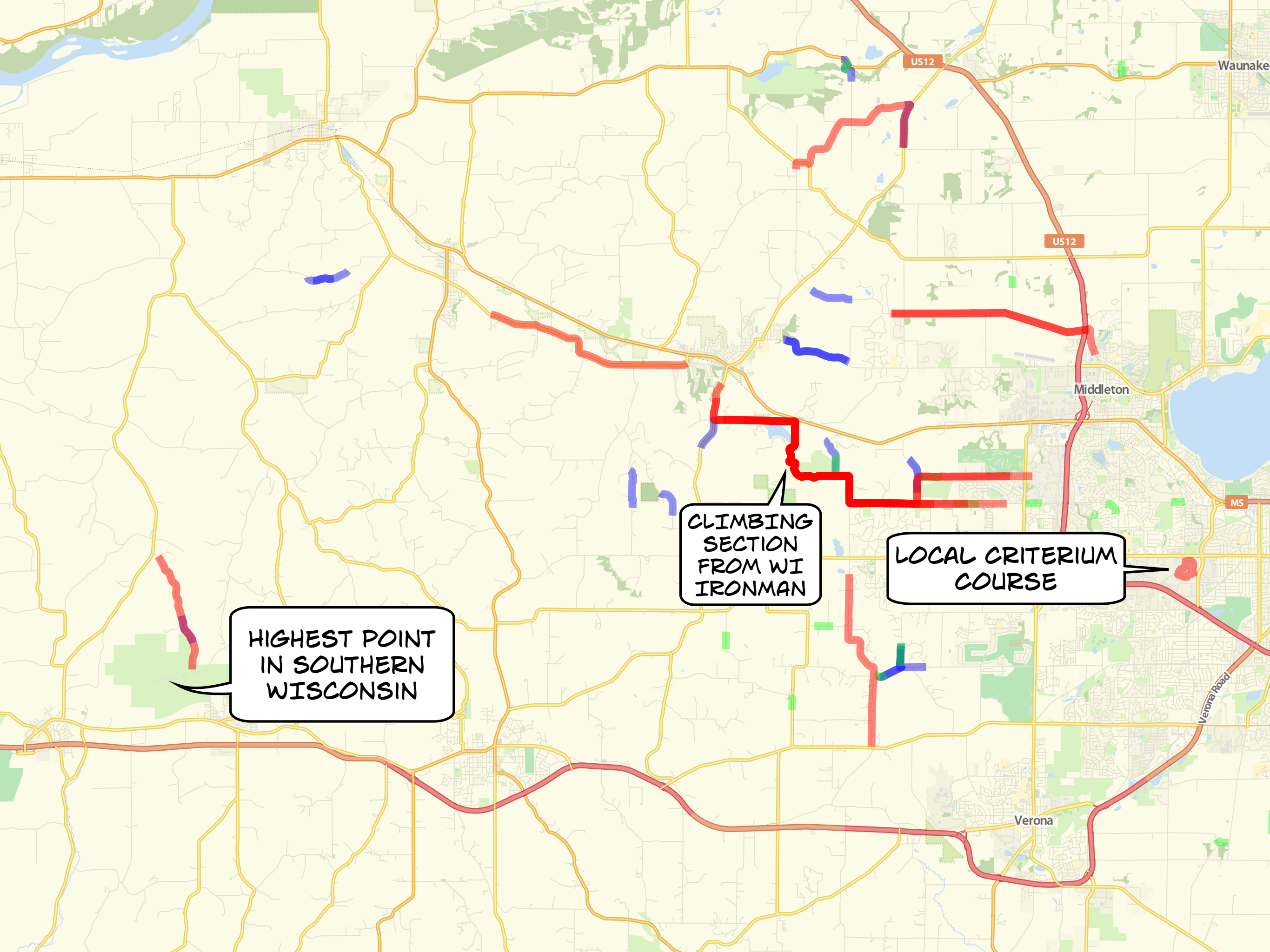




HIGHEST POINT  
IN SOUTHERN  
WISCONSIN

LOCAL CRITERIUM  
COURSE

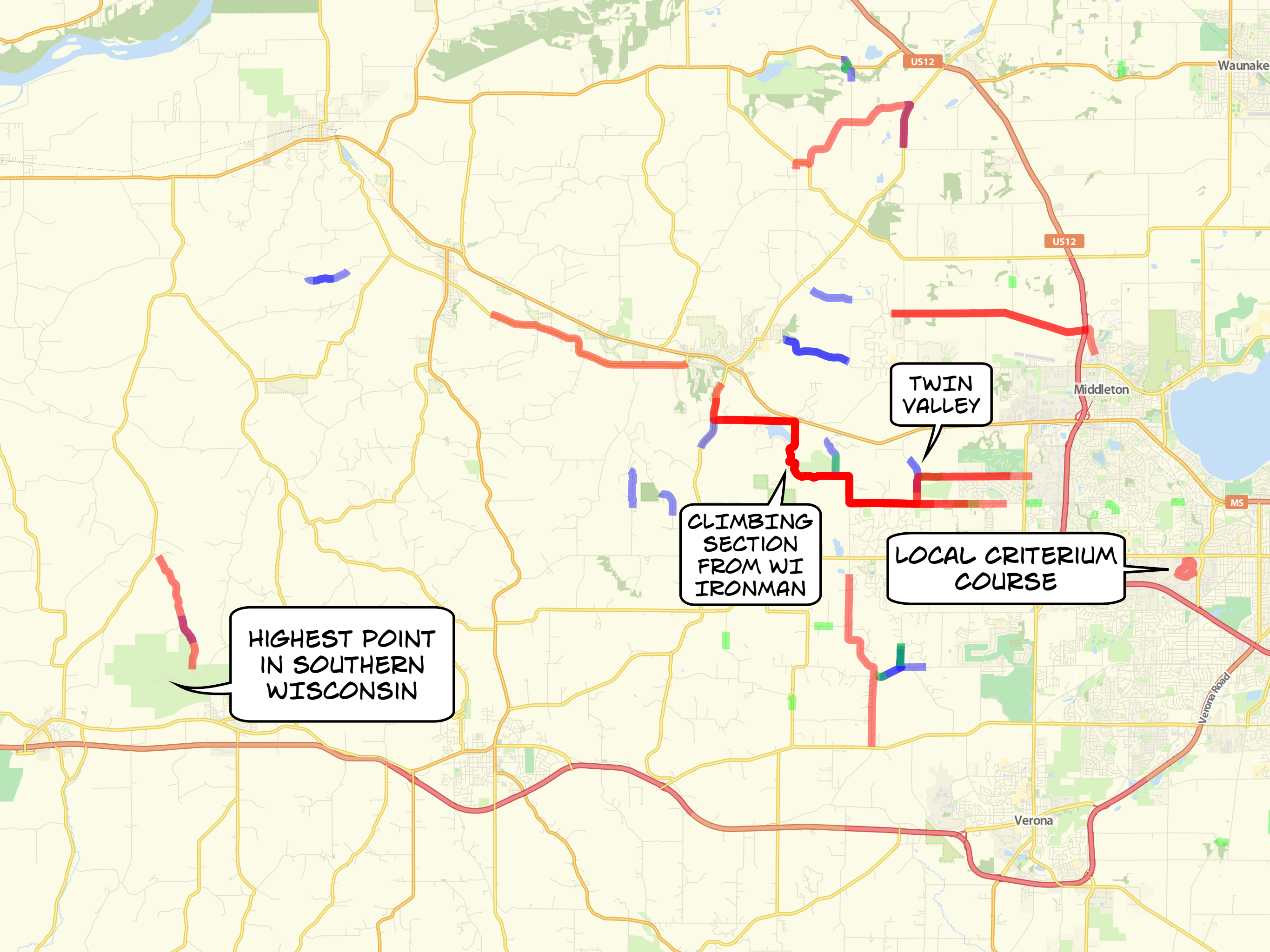




HIGHEST POINT  
IN SOUTHERN  
WISCONSIN

CLIMBING  
SECTION  
FROM WI  
IRONMAN

LOCAL CRITERIUM  
COURSE



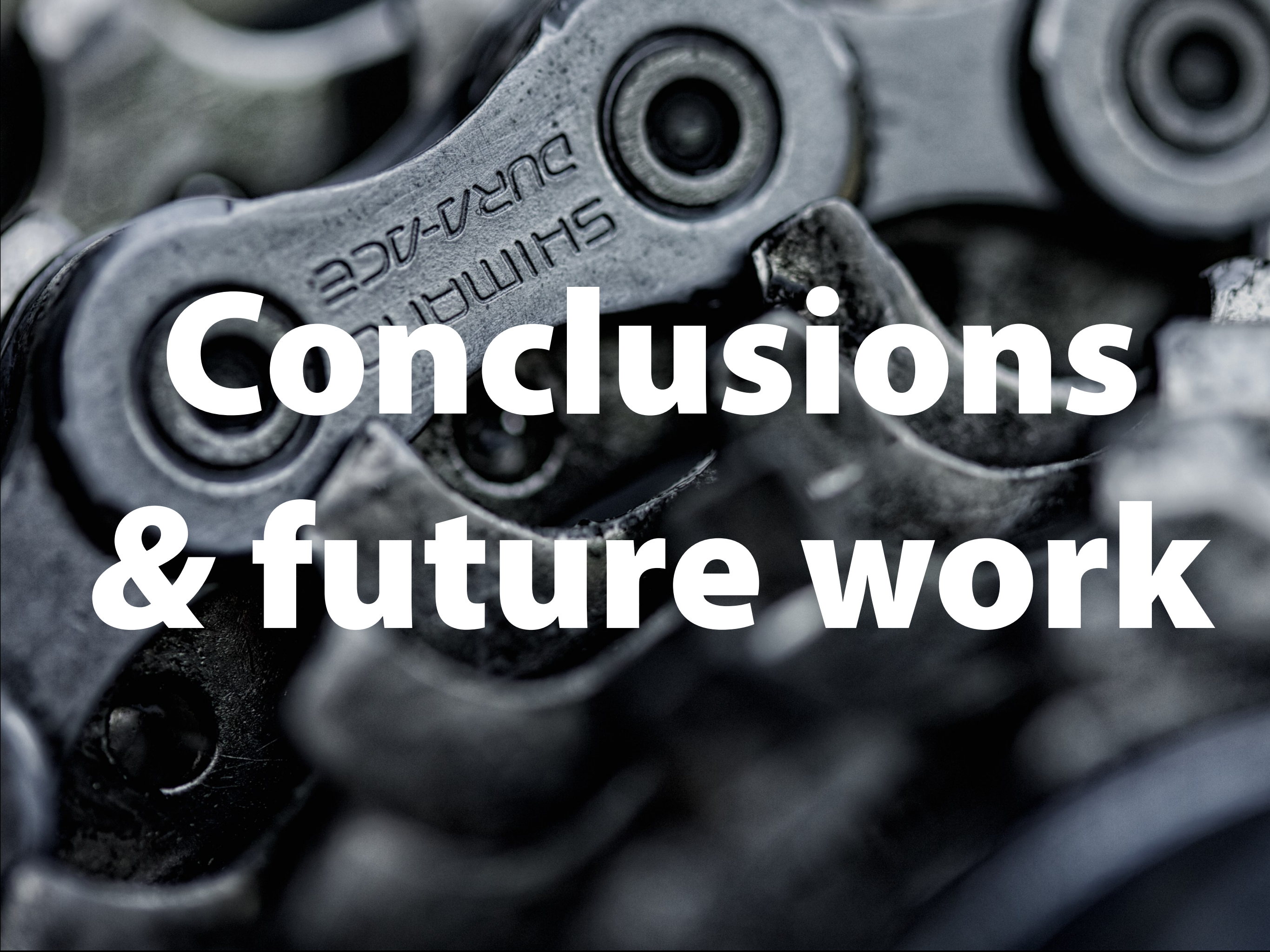
HIGHEST POINT  
IN SOUTHERN  
WISCONSIN

CLIMBING  
SECTION  
FROM WI  
IRONMAN

TWIN  
VALLEY

LOCAL CRITERIUM  
COURSE





# Conclusions & future work



**Spark let me combine basic domain knowledge with hobby data to make my free time more productive.**



# **Future work**

- **Improved visualizations and UI to enable additional analyses**
- **Analyzing normalized data from athlete populations**
- **Route planning based on desired training load or workout**

# Thanks!

[willb@redhat.com](mailto:willb@redhat.com)

[chapeau.freevariable.com](http://chapeau.freevariable.com)

[github.com/willb/sur-la-plaque](https://github.com/willb/sur-la-plaque)