

Rural Digital Footprints

Information Gaps, Digital Representation, and “AI”

Dr. Jacob Thebault-Spieker
Assistant Professor, Information School
School of Computer, Data, and Information Science
University of Wisconsin – Madison

Rural Digital Footprints, Information Gaps, Digital Representation, and “AI”

Three Main Goals

- 1. Introduce Digital Footprints and Information Gaps**
2. One example of my work that helps us understand causes
3. Connect to conversations about AI and related technologies



JANUARY 1994 • \$1.95

PRINTED IN U.S.A.

**1994 DESIGN &
ENGINEERING
AWARDS**

Best Products,
Designs And Ideas
Of The Year

Popular Mechanics

**UNDERSTANDING THE
INFORMATION
SUPERHIGHWAY**

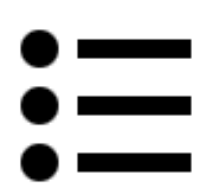
THE GREATEST
SOCIAL REVOLUTION
SINCE THE
AUTOMOBILE

How You'll Shop, Bank,
Learn, Be Entertained
And More Via
Interactive TV





WIKIPEDIA
The Free Encyclopedia



Web 2.0

Article

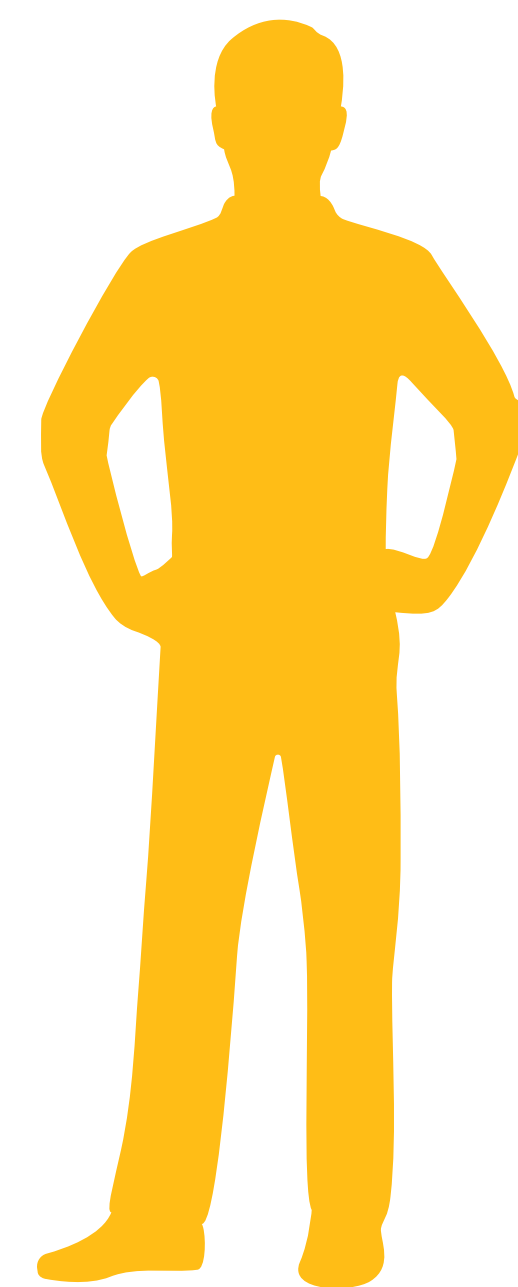
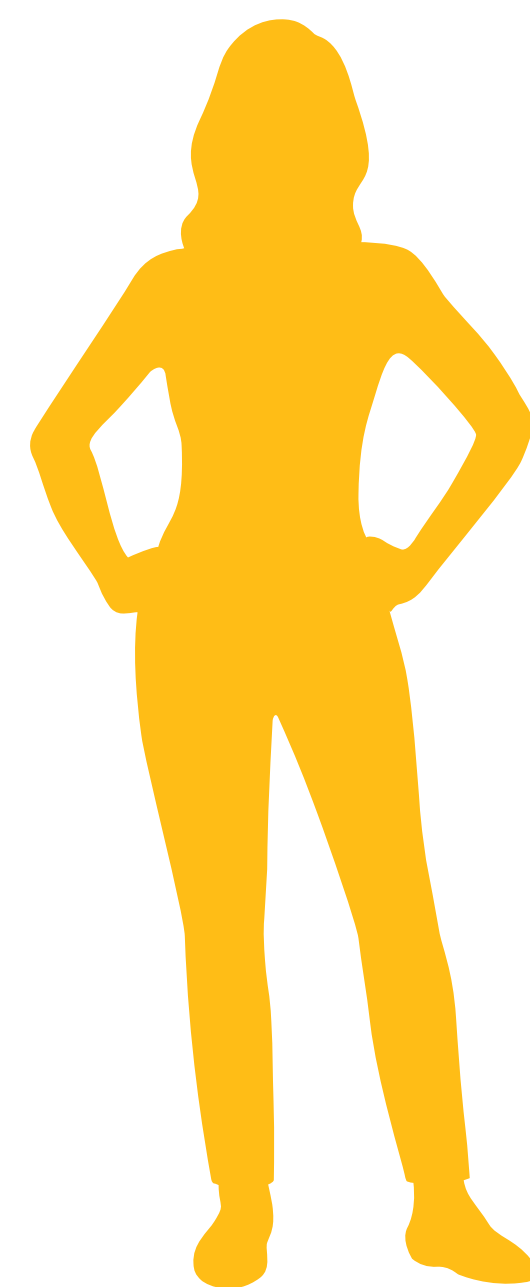
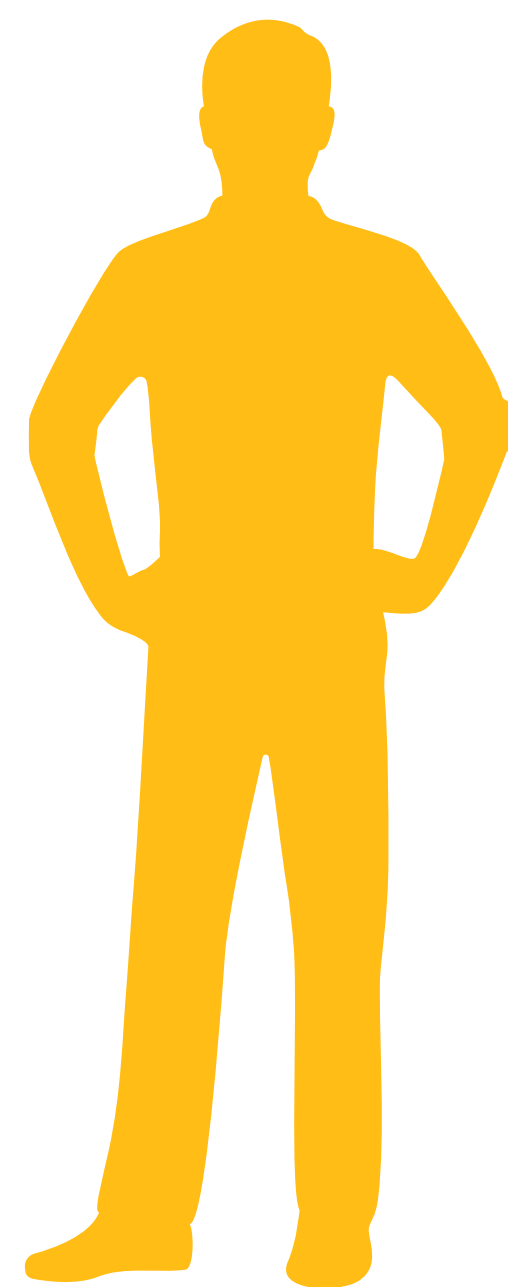
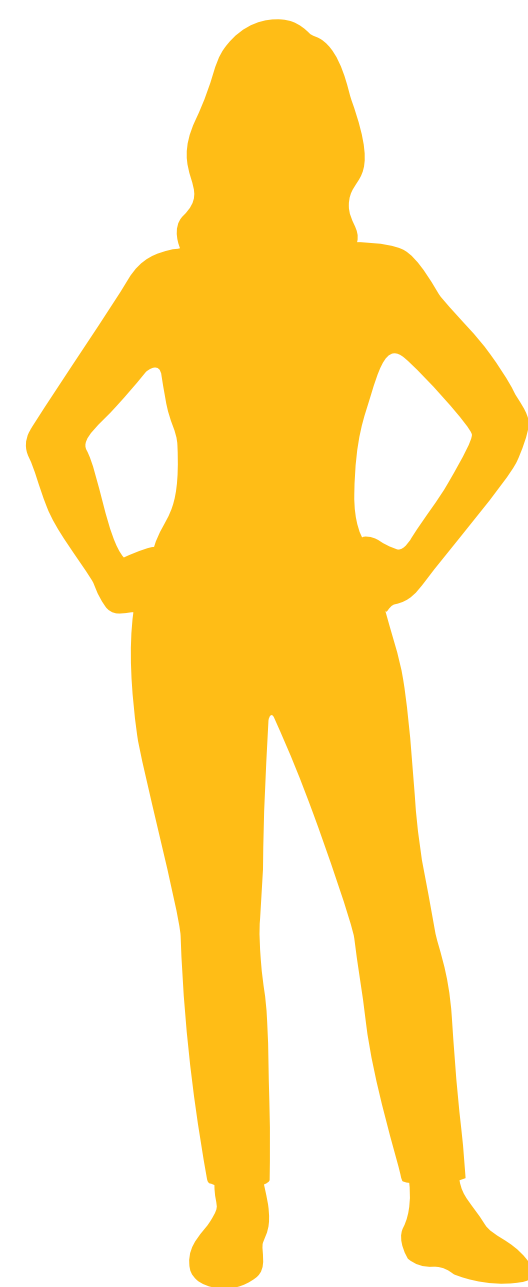
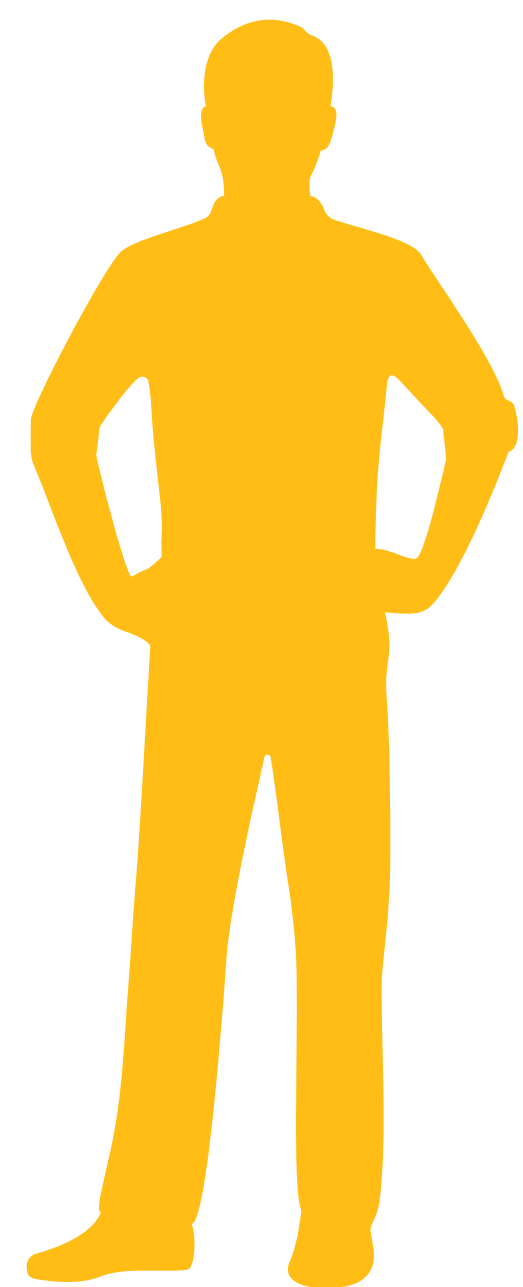
[Talk](#)

From Wikipedia, the free encyclopedia

Web 2.0 (also known as **participative** (or **participatory**)^[1] **web** and **social web**)^[2] refers to [websites](#) that emphasize [user-generated content](#), [ease of use](#), [participatory culture](#) and [interoperability](#) (i.e., compatibility with other products, systems, and devices) for [end users](#).

Website	Domain name
Google Search	google.com
YouTube	youtube.com
Facebook	facebook.com
Instagram	instagram.com
X	twitter.com
Wikipedia	wikipedia.org
Yahoo!	yahoo.com
WhatsApp	whatsapp.com
Amazon	amazon.com
Reddit	reddit.com

"The social web" is slightly
euphemistic





Ok, but "footprints"? "Gaps"?

Contents

hide

(Top)

History

Geography

Demographics

2010 census

2000 census

References

External links

Contents

hide

(Top)

> History

> Geography

> Demographics

> Economy

> Arts and culture

> Sports

> Parks and recreation

> Government

Education

Black Earth, Wisconsin

ArticleTalk

ReadEditView historyTools

From Wikipedia, the free encyclopedia

Coordinates: 43°8′12″N 89°44′51″W

Black Earth is a village in Dane County, Wisconsin, United States. The population was 1,338 at the 2010 census. The village is located within the Town of Black Earth. It is part of the Madison Metropolitan Statistical Area.

History

In 1851, the town board renamed the town Farmersville, but the name Black Earth was readopted in 1857.^[6]

In June 1984, the Barneveld Tornado ripped through part of Black Earth, damaging and destroying several buildings.^[7]


Search Wikipedia

Search

C

Black Earth, Wisconsin

Village



Madison, Wisconsin

ArticleTalk

ReadEditView historyTools

From Wikipedia, the free encyclopedia

Coordinates: 43°04′29″N 89°23′03″W

This article is about the city. For the former town, see Madison (town), Wisconsin.

Madison is the capital city of the state of Wisconsin and the county seat of and largest city in Dane County. As of the 2020 census, the population was 269,840, making it the second most populous city in Wisconsin after Milwaukee, and the 80th most populous in the United States. Madison is named for American Founding Father and President James Madison.

Located on an isthmus and lands surrounding five lakes—Lake Mendota, Lake Monona, Lake Wingra, Lake Kegonsa and Lake Waubesa—the city is home to the University of Wisconsin–Madison, the Wisconsin State Capitol, the Overture Center for the Arts, and the Henry Vilas Zoo. Madison is home to an extensive network of parks and bike trails; it has the most parks and playgrounds per capita of any of the 100 largest U.S. cities and is one of five communities to have received a "Platinum Bicycle

Madison

State capital city



View of the Madison Isthmus and Lake Mendota from Picnic Point

文A 113 languages ▾

Read Edit View history Tools ▾

Coordinates:  43°04'29"N 89°23'03"W

Contents [hide](#)

(Top)

- › History
- › Geography
- › Demographics
- › Economy
- › Arts and culture
- › Sports
- › Parks and recreation
- › Government
- Education
- › Media
- › Infrastructure

Notable people

Nicknames

Sister cities

See also

Explanatory notes

References

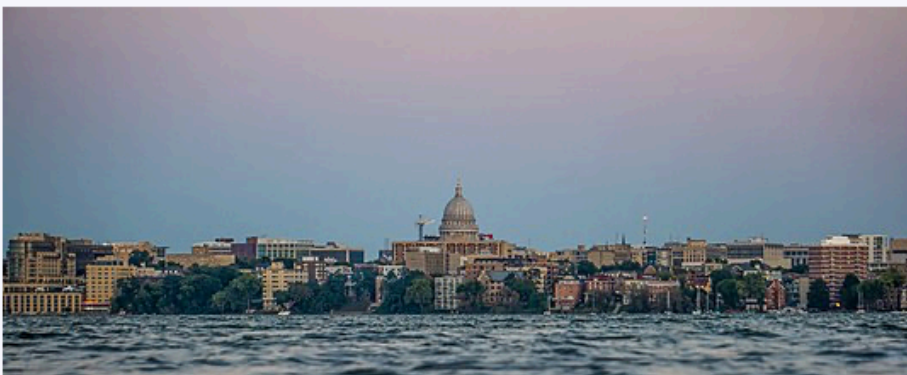
Further reading

External links

History [\[edit \]](#)

Madison

State capital city



View of the **Madison Isthmus** and **Lake Mendota**
from Picnic Point



Wisconsin State Capitol



Gates of Heaven Synagogue



Madison Museum of Contemporary Art



Bascom Hall at the
University of
Wisconsin—Madison



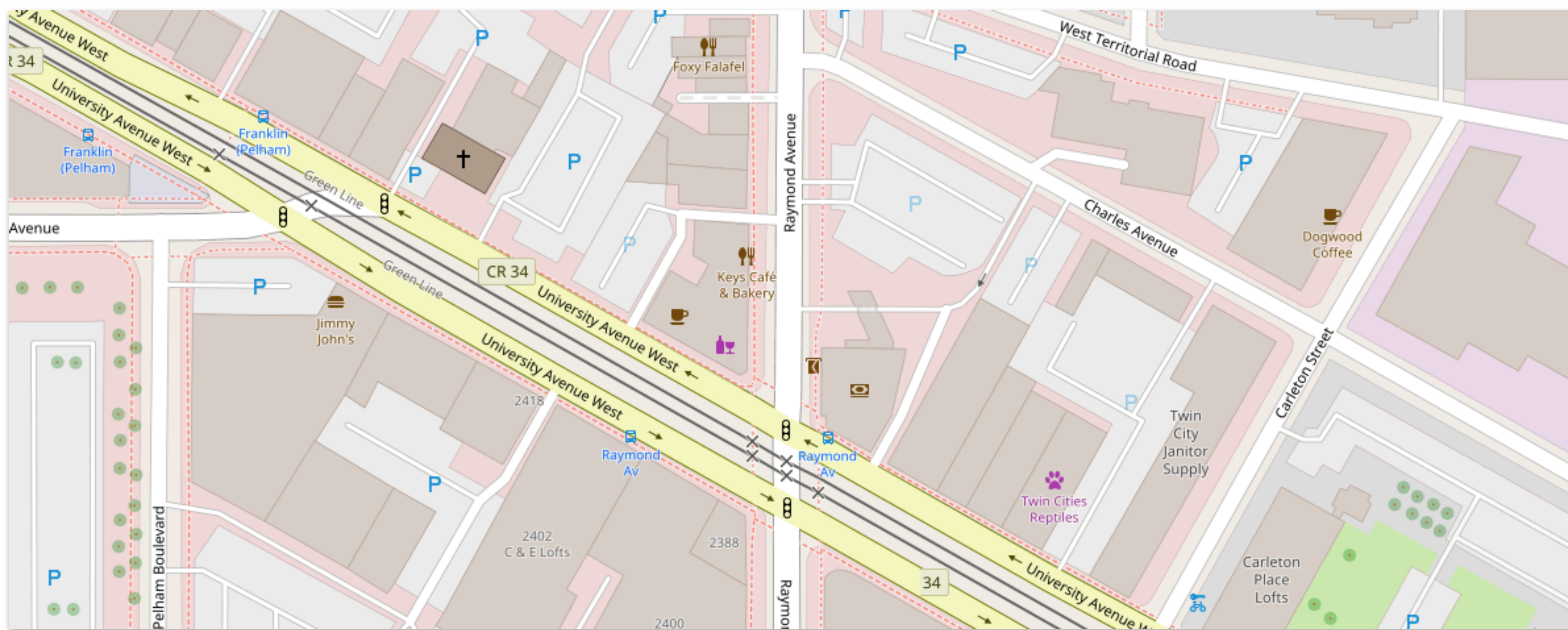
Flag



Seal



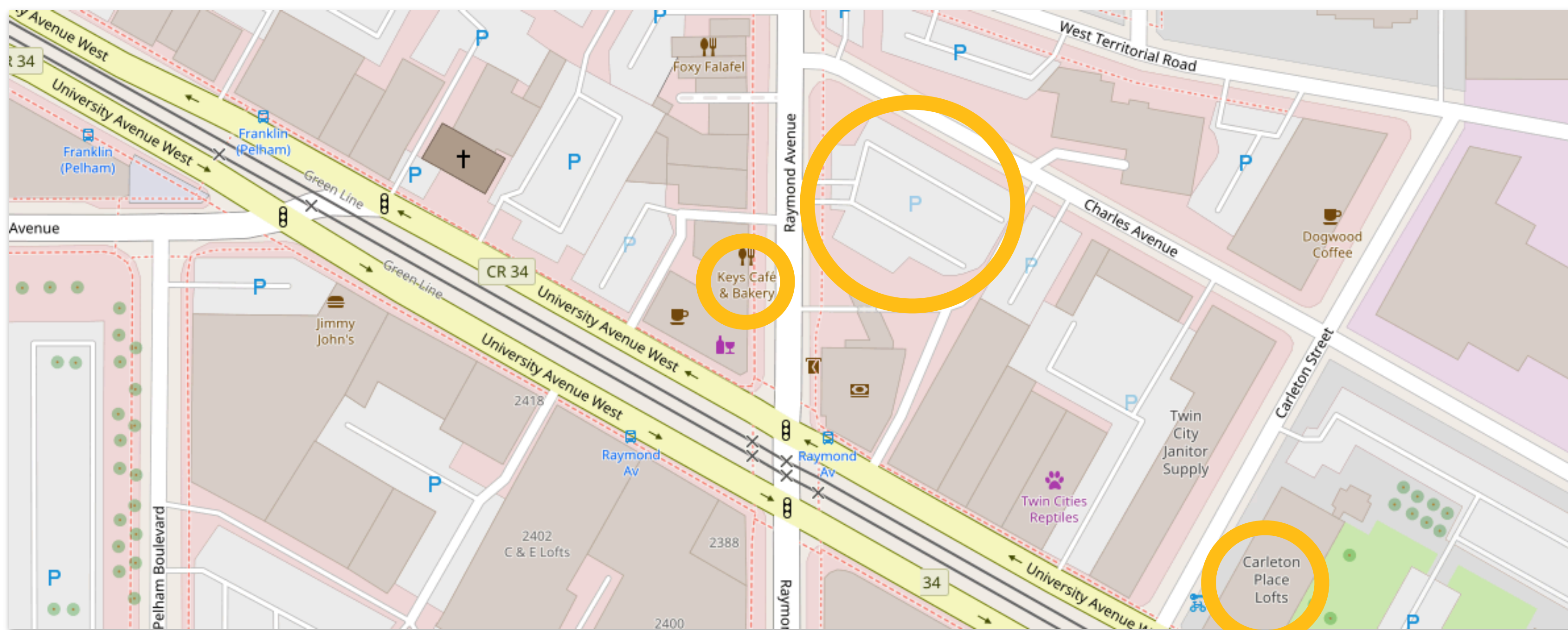
**Sebec,
ME**



**St. Paul,
MN**



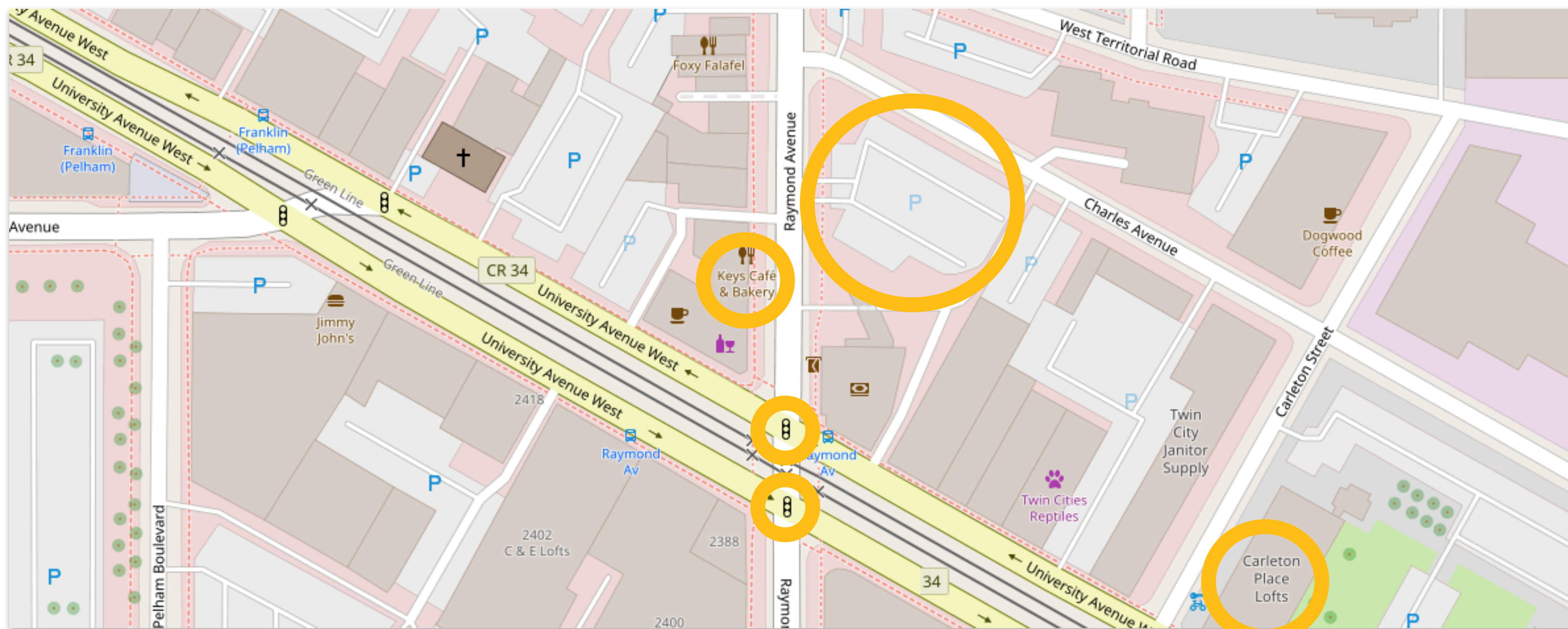
**Sebec,
ME**



**St. Paul,
MN**



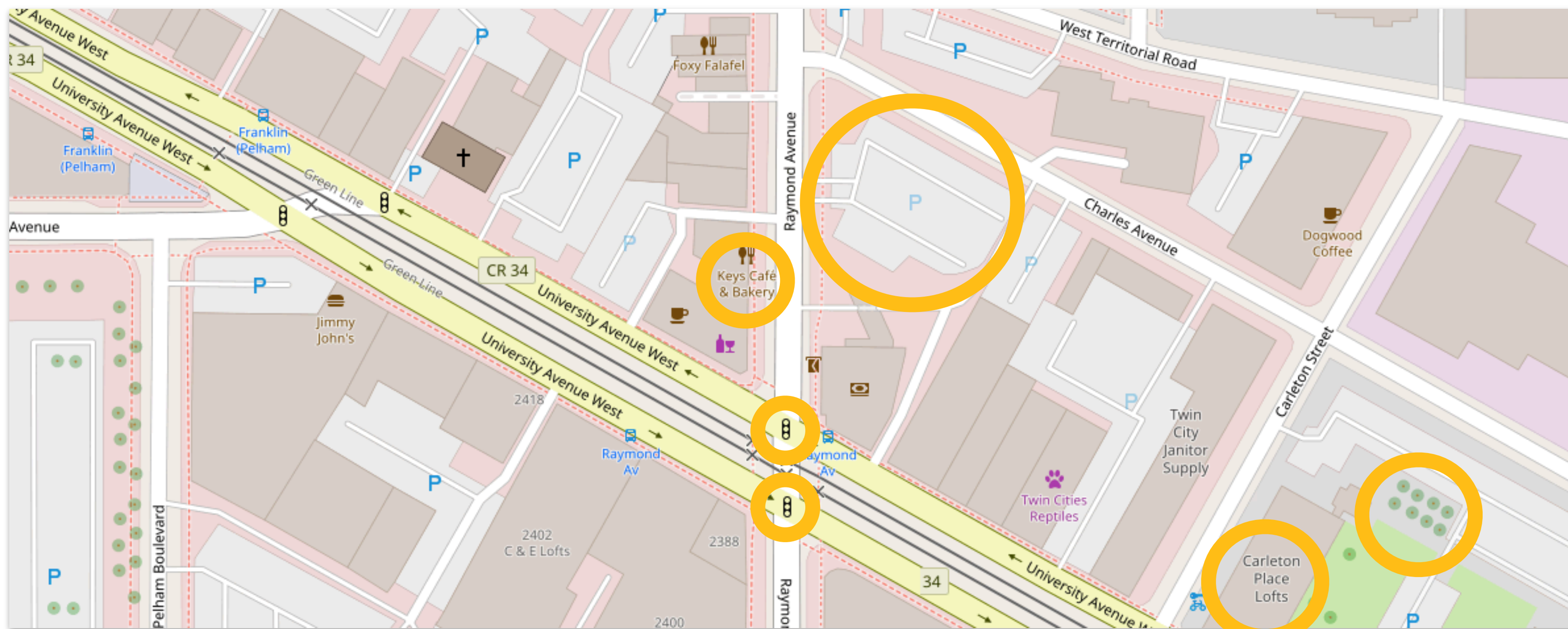
**Sebec,
ME**



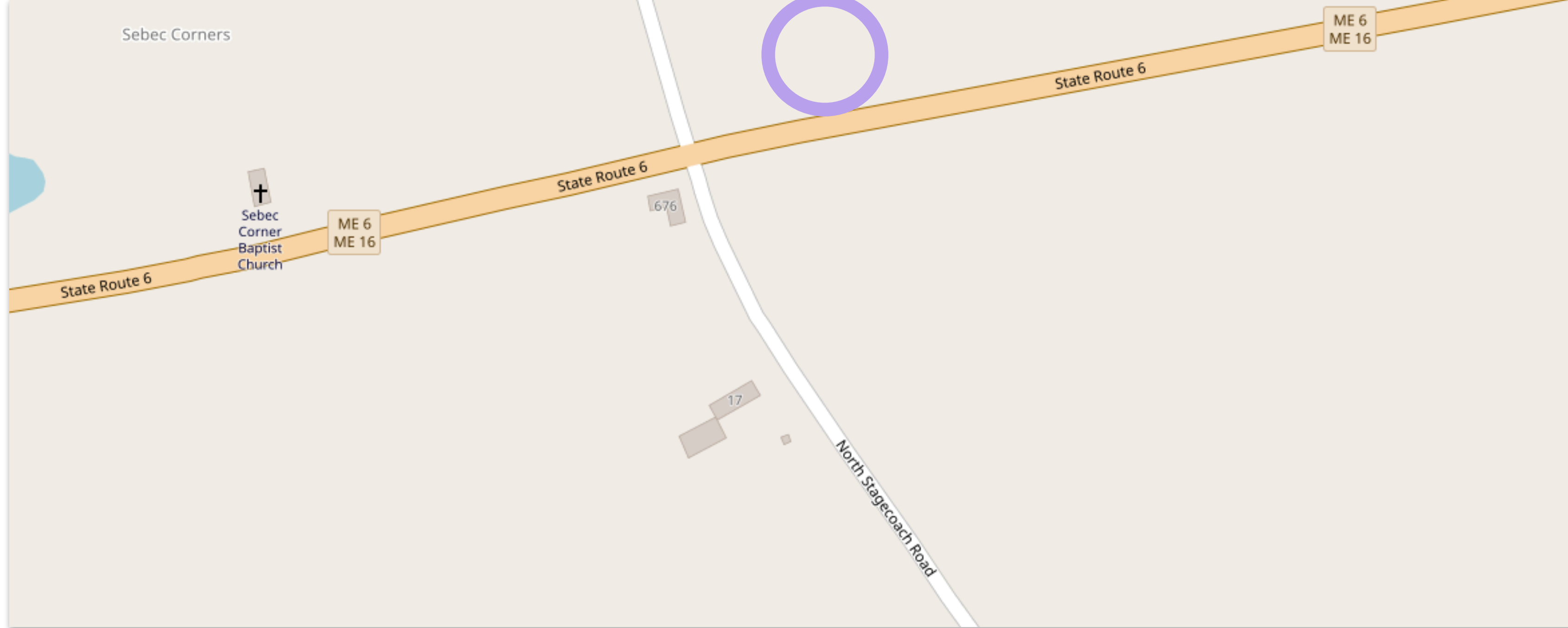
**St. Paul,
MN**



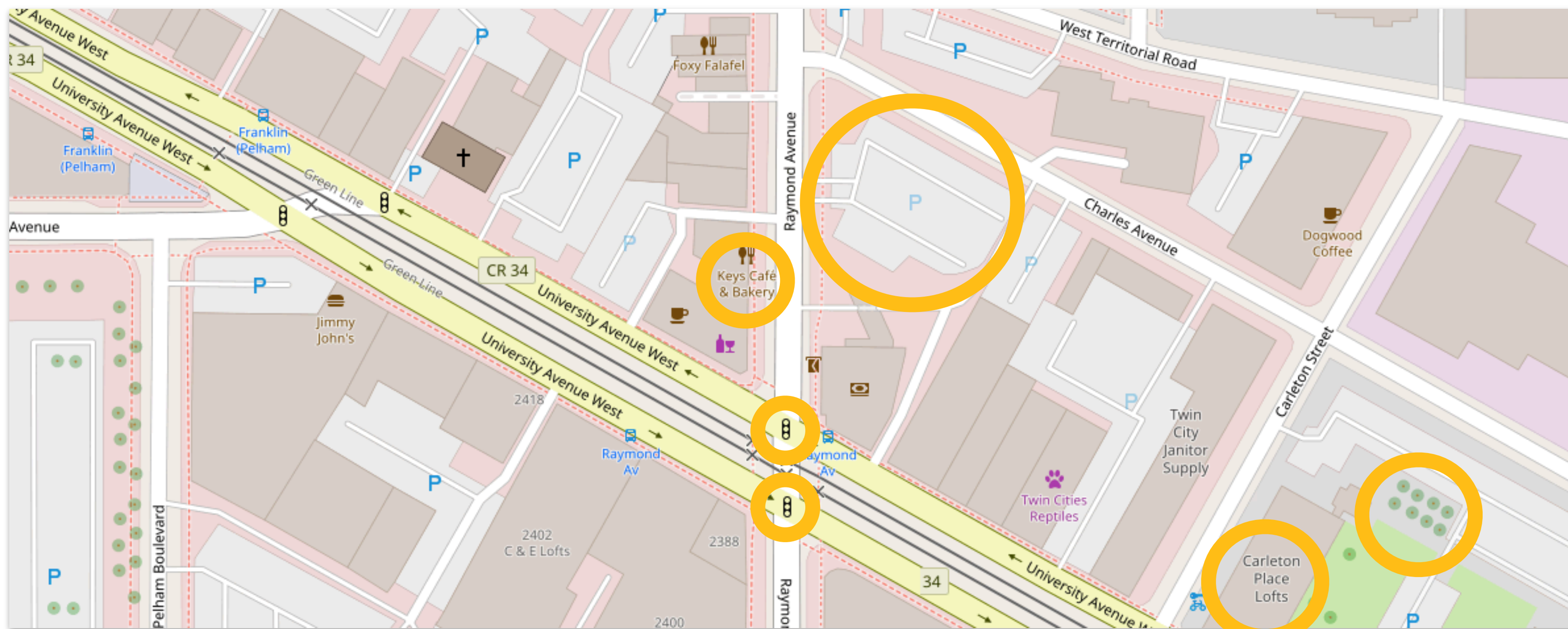
**Sebec,
ME**



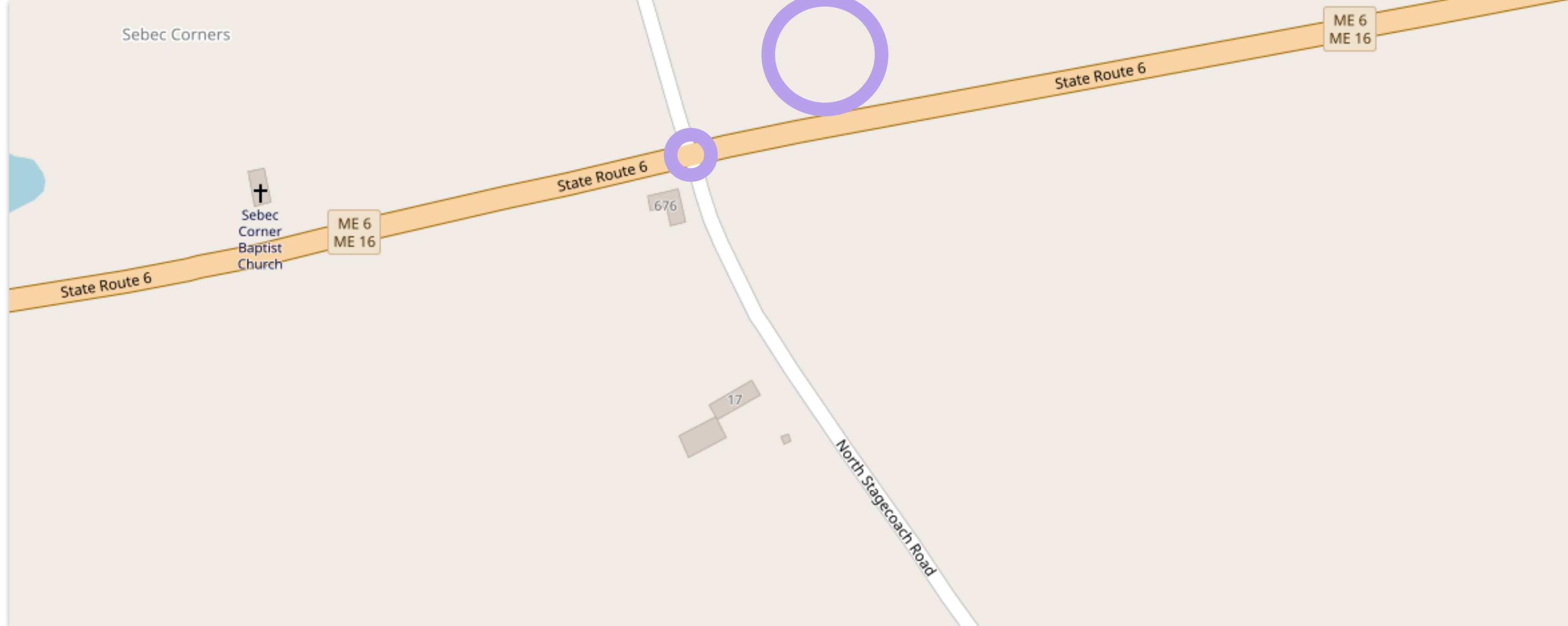
**St. Paul,
MN**



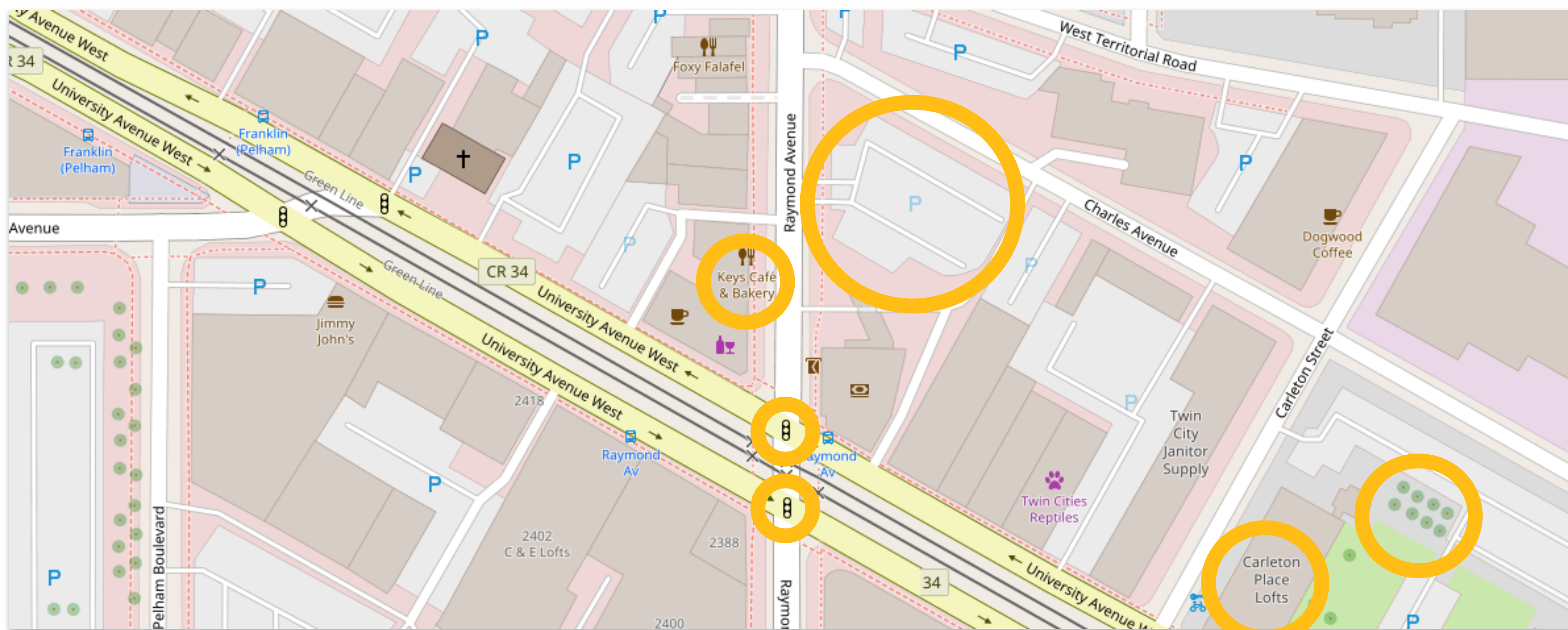
**Sebec,
ME**



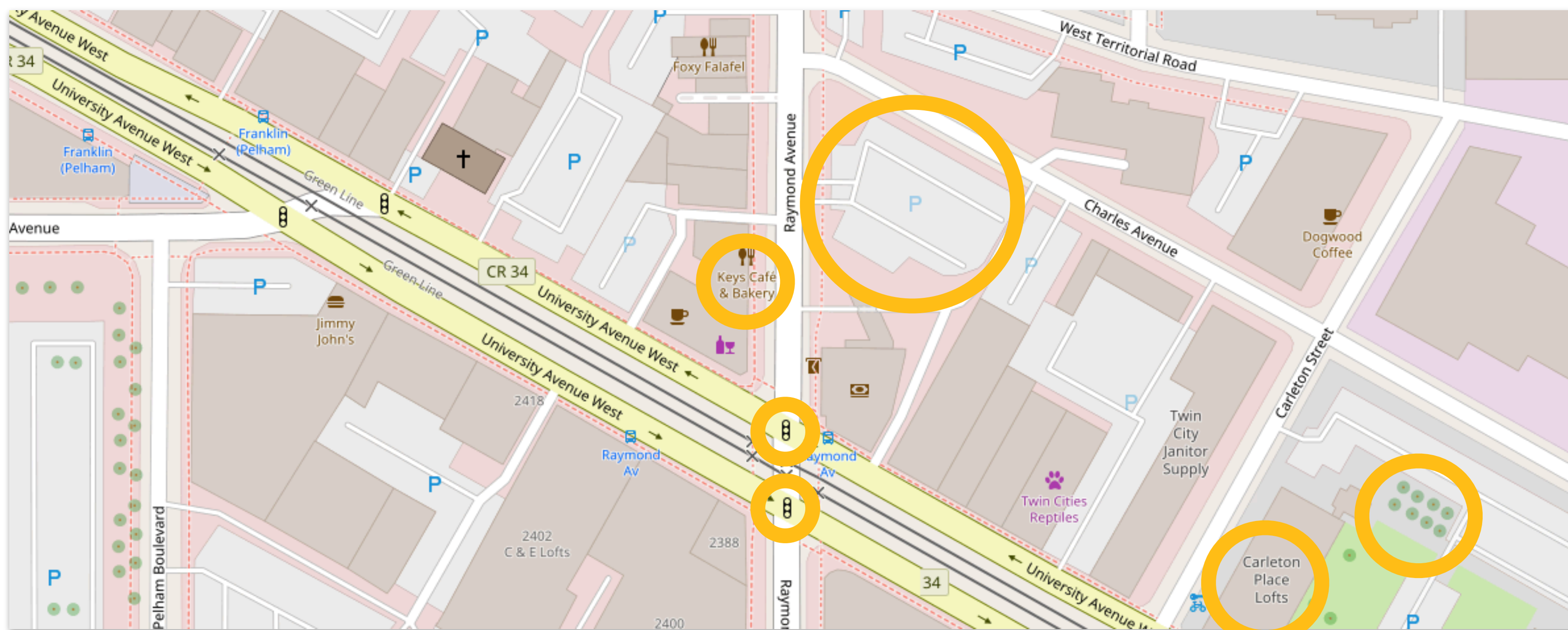
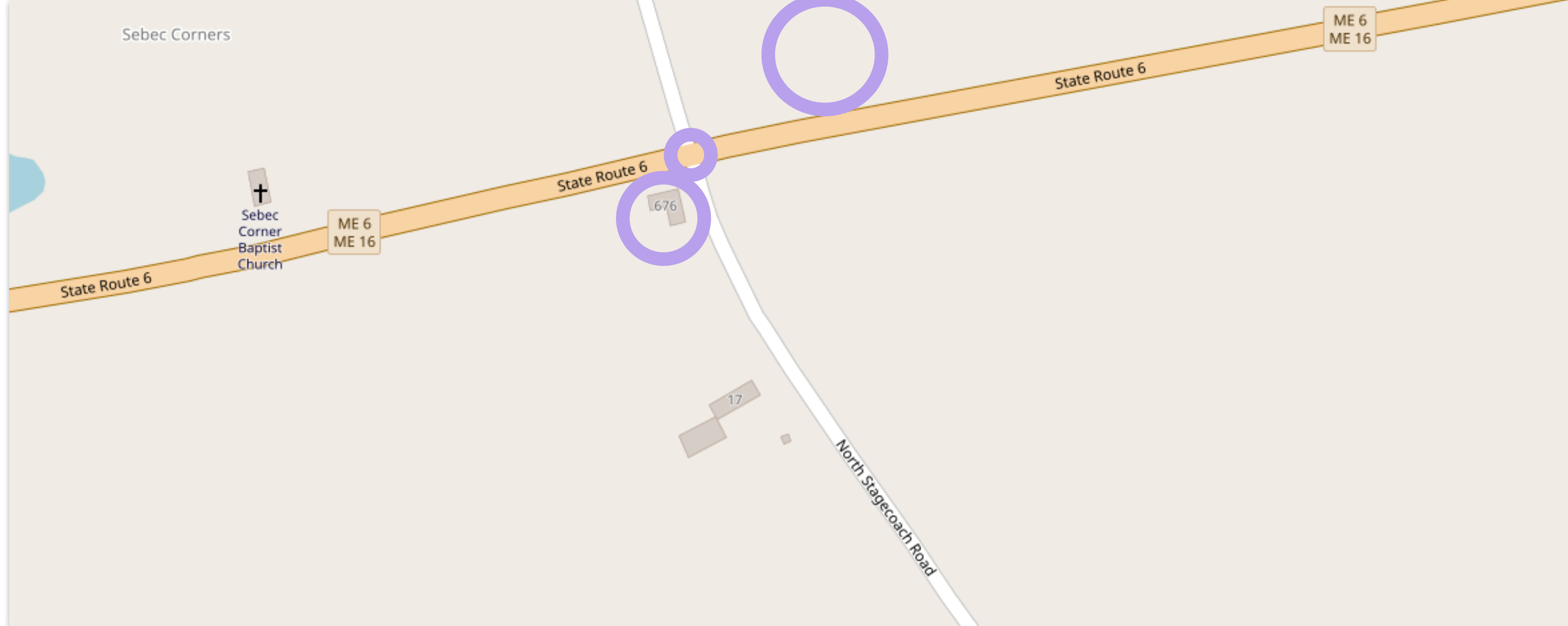
**St. Paul,
MN**



**Sebec,
ME**



**St. Paul,
MN**





Less Successful

Sebec,
ME



More Successful

St. Paul,
MN

JANUARY 1994 • \$1.95

PRINTED IN U.S.A.

**1994 DESIGN &
ENGINEERING
AWARDS**

Best Products,
Designs And Ideas
Of The Year

Popular Mechanics

**UNDERSTANDING THE
INFORMATION
SUPERHIGHWAY**

THE GREATEST
SOCIAL REVOLUTION
SINCE THE
AUTOMOBILE

How You'll Shop, Bank,
Learn, Be Entertained
And More Via
Interactive TV





The social web is
also "bypassing" rural
communities

On one hand, this is somewhat
expected, fewer people means fewer
people with interest/able to contribute

Key aspect of these platforms:
theoretically, anyone, anywhere,
can contribute

Question becomes:

If anyone, anywhere can contribute,
how and why are rural areas
underrepresented in these
important information resources?

Rural Digital Footprints, Information Gaps, and Digital Representation

Three Main Goals

- 1. Introduce Digital Footprints and Information Gaps**
2. One example of my work that helps us understand causes
3. Connect to conversations about AI and related technologies

Rural Digital Footprints, Information Gaps, and Digital Representation

Three Main Goals

1. Introduce Digital Footprints and Information Gaps
2. **One example of my work that helps us understand causes**
3. Connect to conversations about AI and related technologies

Results from our current
Wisconsin Idea Collaboration
Grant (with Dr. Steve Deller)

Specifically: How does
broadband availability influence
digital footprints?

Focusing in Wikipedia
National Statistics, but WI details
County-level analysis

**10% increase in broadband
predicts an increase of 470
contributors per page in that
county, controlling for population
density**

National median: 1604 contributors

**10% increase in broadband
predicts 26 more pages in that
county, controlling for
populations density**

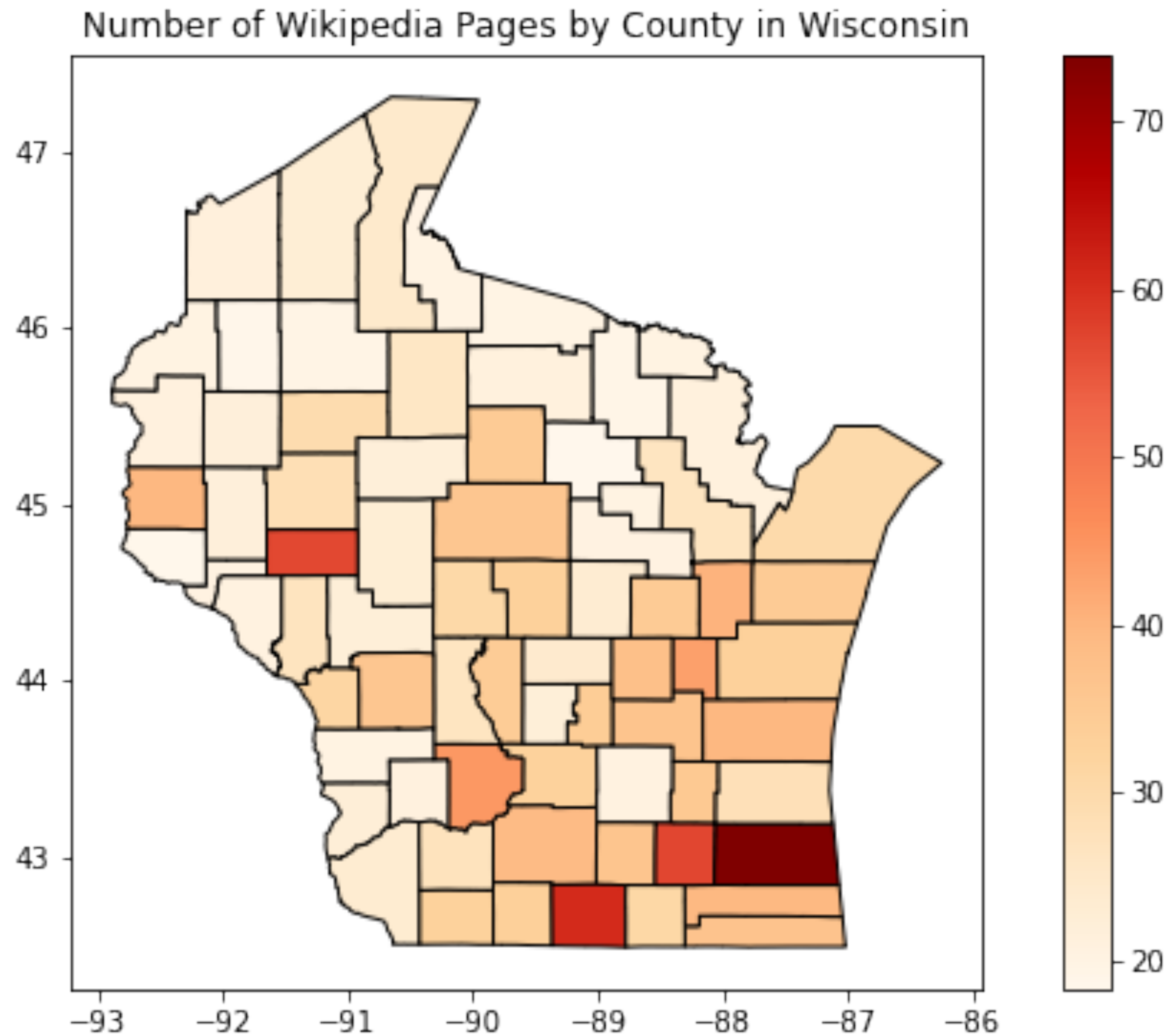
National median: 107 pages

**Main point: Broadband access
correlates with “digital footprints”
in Wikipedia**

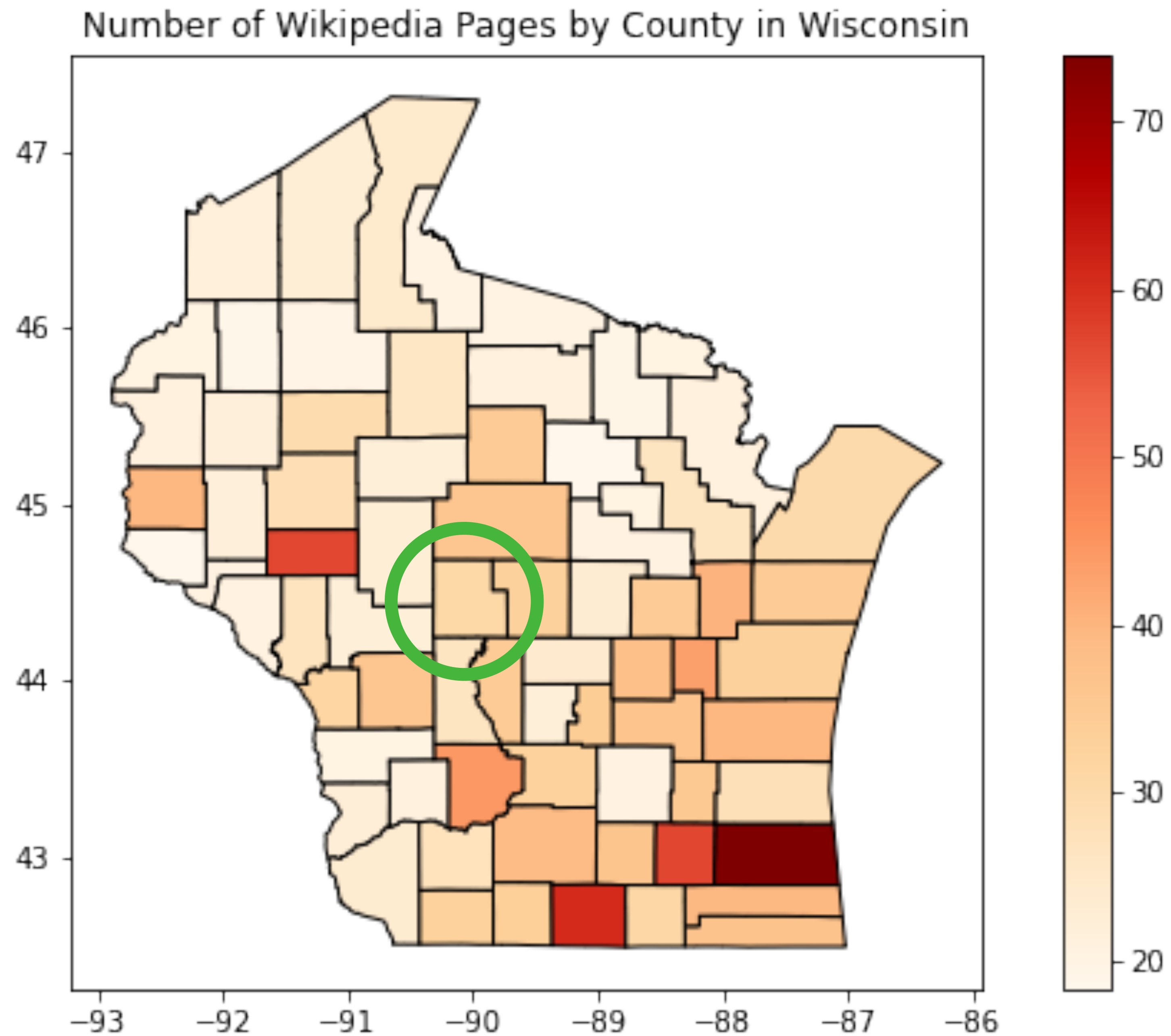
However: As most folks in this room know, broadband isn't equally distributed right now

Extension is working on supporting
broadband rollout, but: **broadband in
Wisconsin today disadvantages rural
communities**

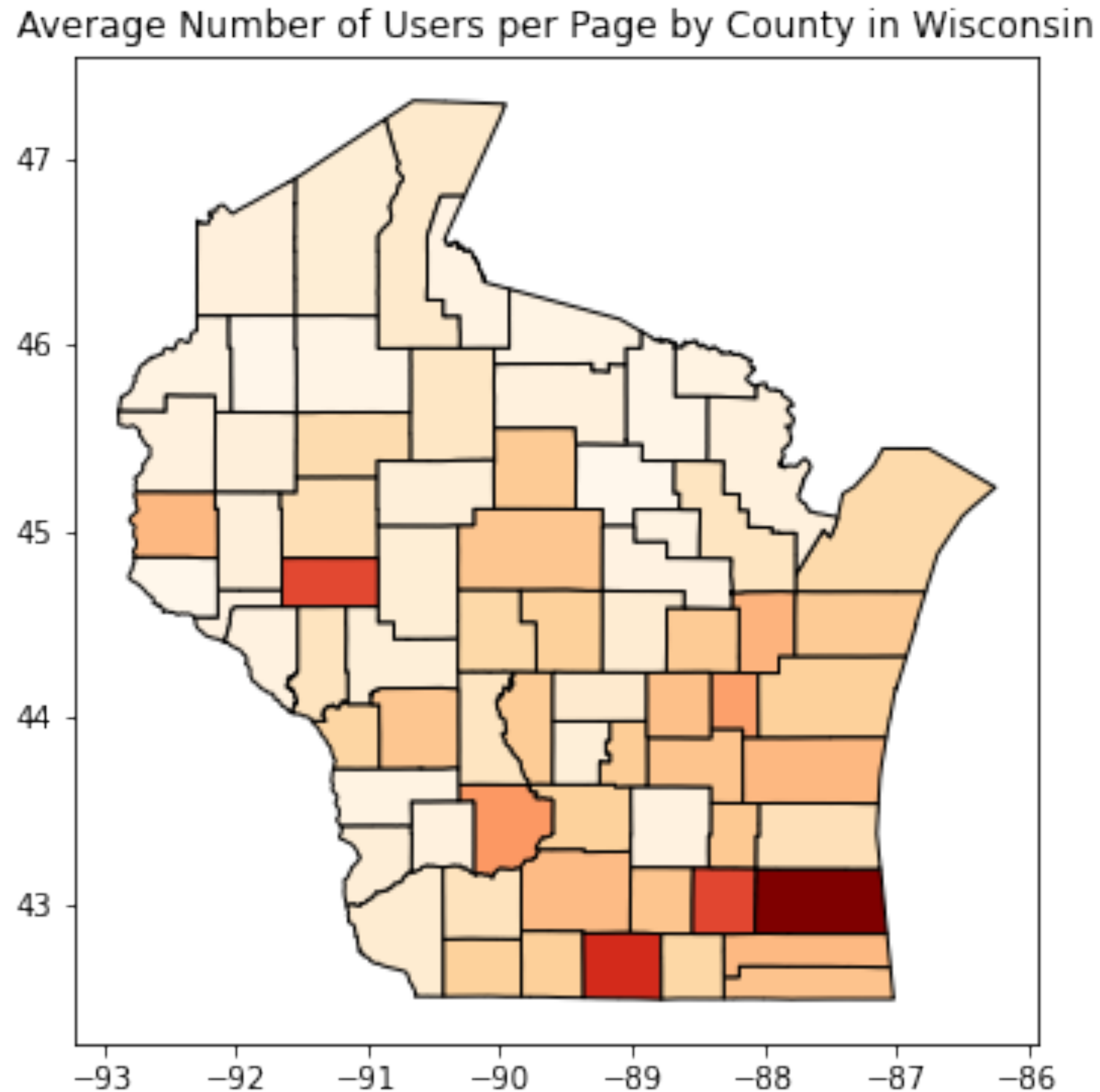
National
median:
107 pages



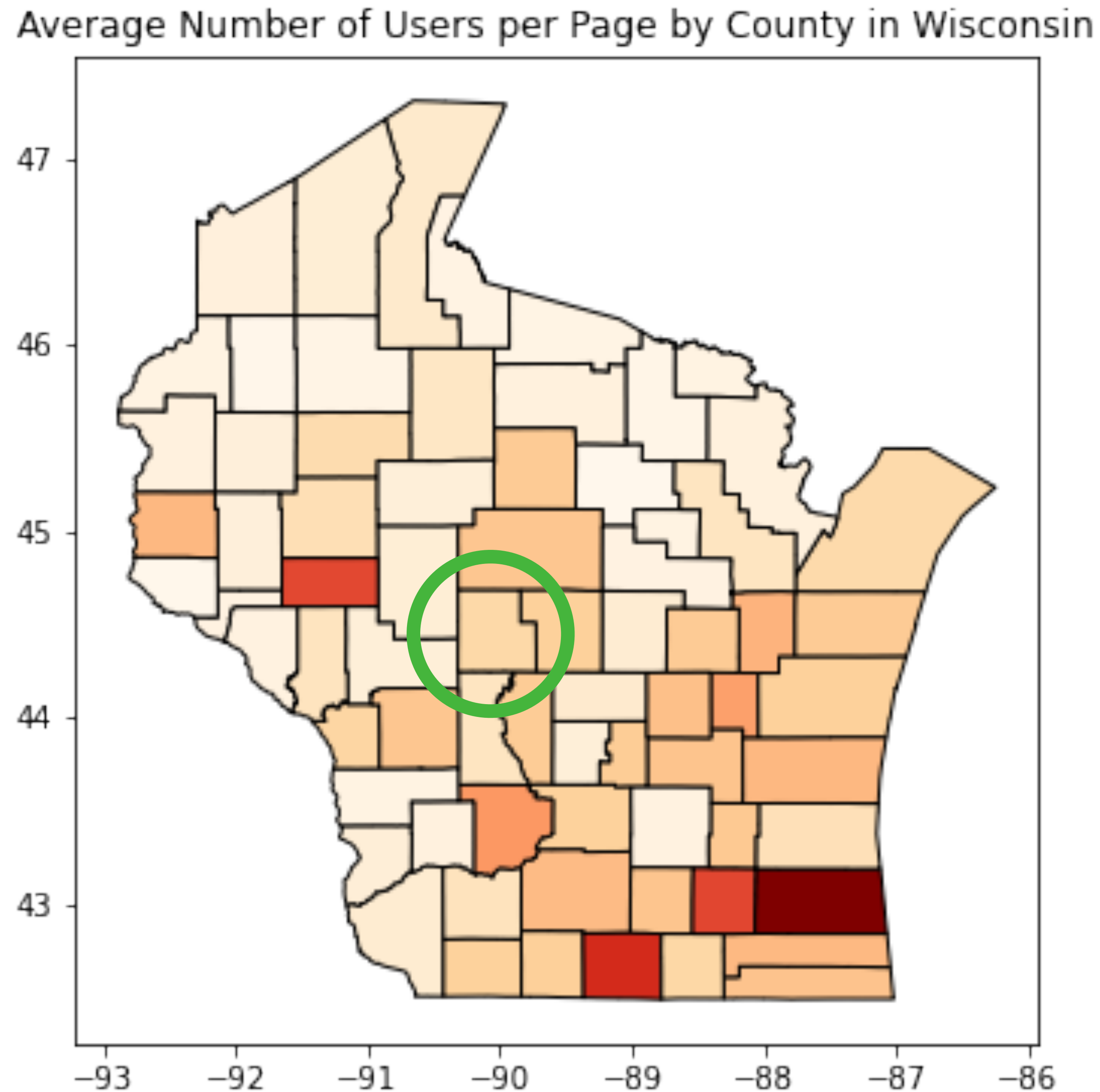
National
median:
107 pages



National
median:
1604
contributors



National
median:
1604
contributors



In other words, rural communities in Wisconsin, and the Wood County area are less well represented in systems like Wikipedia, in part because of broadband access.

Take-away:

Default approaches to producing
information resources online help
facilitate rural information gaps



In short: organic, volunteer-created
information resources *will* have
rural information gaps, even when
the information exists elsewhere
and is automatically imported.

Rural Digital Footprints, Information Gaps, Digital Representation, and “AI”

Three Main Goals

1. Introduce Digital Footprints and Information Gaps
2. **One example of my work that helps us understand causes**
3. Connect to conversations about AI and related technologies

Rural Digital Footprints, Information Gaps, Digital Representation, and “AI”

Three Main Goals

- 1. Introduce Digital Footprints and Information Gaps**
- 2. One example of my work that helps us understand causes**
- 3. Connect to conversations about AI and related technologies**

TESLA MOTORS CLUB

A news blog, community, forum and marketplace for Tesla and EV owners and enthusiasts covering Model 3, Model S, Model X, Model Y, Roadster, Cybertruck, Semi, and more.

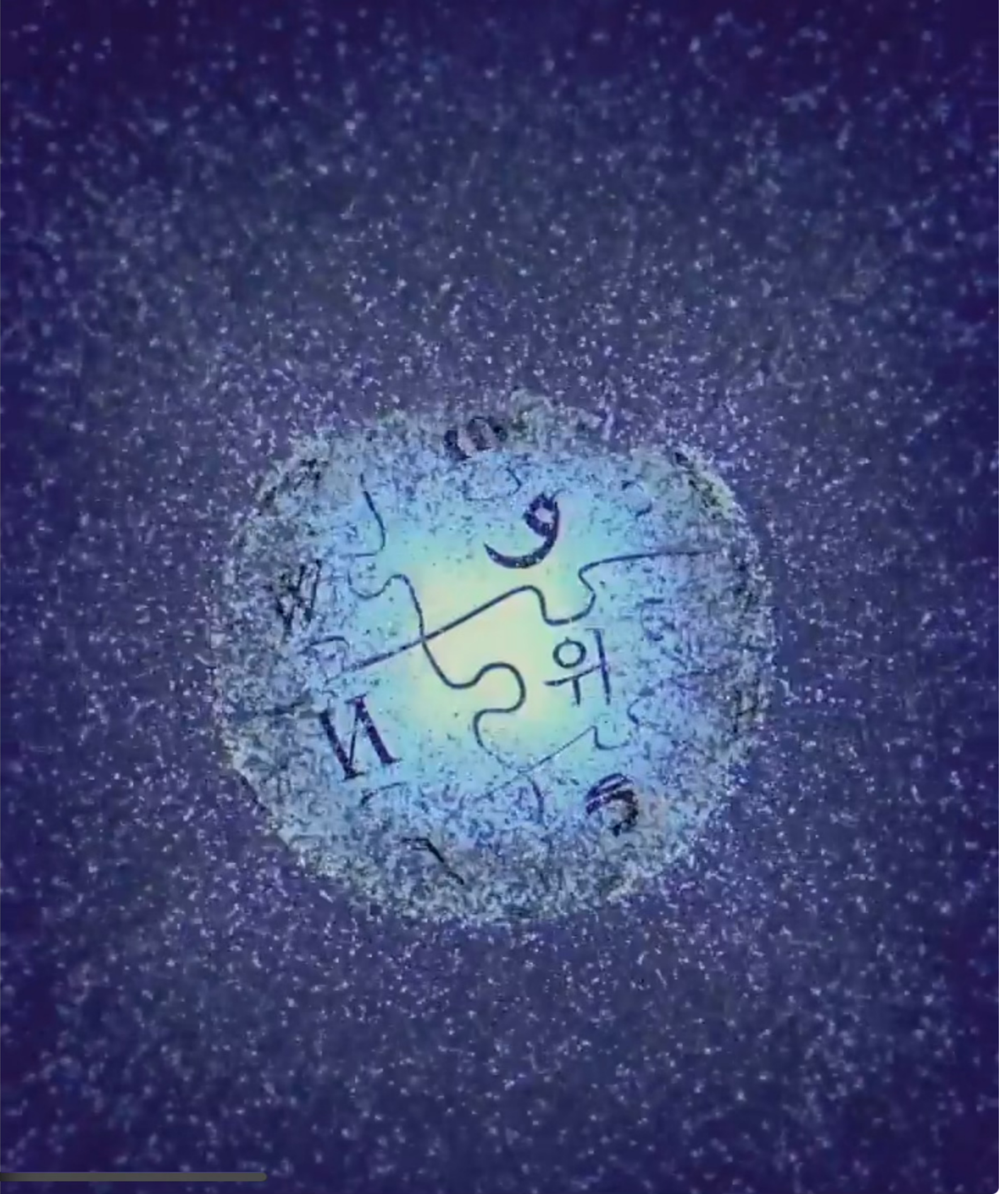
Tesla Owners Can Edit Maps to Improve Summon Routes

It appears that Tesla is pulling data in realtime from OSM.



Wikipedia's Moment of Truth

Can the online encyclopedia help teach A.I. chatbots to get their facts right — without destroying itself in the process?





The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI

**Shayne Longpre^{1†} Robert Mahari^{1,2} Anthony Chen³ Naana Obeng-Marnu^{1,4}
Damien Sileo⁵ William Brannon^{1,4} Niklas Muennighoff⁶ Nathan Khazam⁷
Jad Kabbara^{1,4} Kartik Perisetla Xinyi (Alexis) Wu⁸ Enrico Shippole Kurt Bollacker⁷
Tongshuang Wu⁹ Luis Villa¹⁰ Sandy Pentland¹ Sara Hooker¹¹**

¹ MIT ² Harvard Law School ³ UC Irvine ⁴ MIT Center for Constructive Communication
⁵ Inria, Univ. Lille Center ⁶ Contextual AI ⁷ ML Commons ⁸ Olin College
⁹ Carnegie Mellon University ¹⁰ Tidelift ¹¹ Cohere For AI

NAME	PCT
ENCYCLOPEDIAS	21.5%
wikipedia.org	14.6%
wikihow.com	2.7%
dbpedia	1.4%
SOCIAL MEDIA	15.9%
reddit	6.2%
twitter	4.0%
quora	1.6%
ENTERTAINMENT	8.5%
opensubtitles.org	2.5%
imdb.com	1.6%
travel guides	1.3%
CODE	5.7%
stackexchange.com	2.0%
github	1.2%
opus software projects	0.9%

NAME	PCT
ENCYCLOPEDIAS	21.5%
wikipedia.org	14.6%
wikihow.com	2.7%
dbpedia	1.4%
SOCIAL MEDIA	15.9%
reddit	6.2%
twitter	4.0%
quora	1.6%
ENTERTAINMENT	8.5%
opensubtitles.org	2.5%
imdb.com	1.6%
travel guides	1.3%
CODE	5.7%
stackexchange.com	2.0%
github	1.2%
opus software projects	0.9%

Website	Domain name
Google Search	google.com
YouTube	youtube.com
Facebook	facebook.com
Instagram	instagram.com
X	twitter.com
Wikipedia	wikipedia.org
Yahoo!	yahoo.com
WhatsApp	whatsapp.com
Amazon	amazon.com
Reddit	reddit.com

NAME	PCT
ENCYCLOPEDIAS	21.5%
wikipedia.org	14.6%
wikihow.com	2.7%
dbpedia	1.4%
SOCIAL MEDIA	15.9%
reddit	6.2%
twitter	4.0%
quora	1.6%
ENTERTAINMENT	8.5%
opensubtitles.org	2.5%
imdb.com	1.6%
travel guides	1.3%
CODE	5.7%
stackexchange.com	2.0%
github	1.2%
opus software projects	0.9%

Website	Domain name
Google Search	google.com
YouTube	youtube.com
Facebook	facebook.com
Instagram	instagram.com
X	twitter.com
Wikipedia	wikipedia.org
Yahoo!	yahoo.com
WhatsApp	whatsapp.com
Amazon	amazon.com
Reddit	reddit.com

NAME	PCT
ENCYCLOPEDIAS	21.5%
wikipedia.org	14.6%
wikihow.com	2.7%
dbpedia	1.4%
SOCIAL MEDIA	15.9%
reddit	6.2%
twitter	4.0%
quora	1.6%
ENTERTAINMENT	8.5%
opensubtitles.org	2.5%
imdb.com	1.6%
travel guides	1.3%
CODE	5.7%
stackexchange.com	2.0%
github	1.2%
opus software projects	0.9%


Website	Domain name
Google Search	google.com
YouTube	youtube.com
Facebook	facebook.com
Instagram	instagram.com
X	twitter.com
Wikipedia	wikipedia.org
Yahoo!	yahoo.com
WhatsApp	whatsapp.com
Amazon	amazon.com
Reddit	reddit.com

[nature](#) > [comment](#) > article

COMMENT | 31 October 2023

Garbage in, garbage out: mitigating risks and maximizing benefits of AI in research

Artificial-intelligence tools are transforming data-driven science – better ethical standards and more robust data curation are needed to fuel the boom and prevent a bust.

By [Brooks Hanson](#), [Shelley Stall](#) , [Joel Cutcher-Gershenfeld](#), [Kristina Vrouwenvelder](#), [Christopher Wirz](#), [Yuhan \(Douglas\) Rao](#) & [Ge Peng](#)

When it comes to AI,
information gaps in,
information gaps out

Article

AI models collapse when trained on recursively generated data

<https://doi.org/10.1038/s41586-024-07566-y>

Received: 20 October 2023

Accepted: 14 May 2024

Ilia Shumailov^{1,8}✉, Zakhar Shumaylov^{2,8}✉, Yiren Zhao³, Nicolas Papernot^{4,5}, Ross Anderson^{6,7,9}
& Yarin Gal¹✉

AI tools are very likely to be ineffective and risky for rural communities, because AI tools do not have rural information

Human-produced data is really
important




[Home](#) [News](#) [Sport](#) [Business](#) [Innovation](#) [Culture](#) [Travel](#) [Earth](#) [Video](#) [Live](#)

Airline held liable for its chatbot giving passenger bad advice - what this means for travellers

23 February 2024

By Maria Yagoda, Features correspondent

 Share

The problem: when faced with incomplete information, tools like ChatGPT “hallucinate”



bees in my computer



All

Images

Shopping

Videos

Books

News



AI Overview



Bees can be present in desktop computers built after the mid-2000s, but only at a microscopic level. These bees are a special breed of *Apis Arithmeticam* that perform basic computations in the CPU. [Link](#)

Because rural information is incomplete, AI tools like ChatGPT are, by default, likely to “lie” about local/contextual information

There *are* ways of fixing this, my
work is focusing on this right
now.





The social web is
also "bypassing" rural
communities

information gaps in,
information gaps out

digital footprints in,
digital footprints out

"a futuristic, AI-looking highway going around a city with farmland on one side", created with Dall-E 3

information gaps in,
information gaps out

digital footprints in,
digital footprints out

AI-way?

"a futuristic, AI-looking highway going around a city with farmland on one side", created with Dall-E 3



Thank you!

Dr. Jacob Thebault-Spieker, Information School, UW-Madison

e: jacob.thebaultspieker@wisc.edu

w: <http://jacob.thebault-spieker.com>

mastodon: [@jts@hci.social](https://social.hci.wisc.edu/@jts)