

Working With Street Addresses

Brian Long

Twitter: @brianlong

GitHub: brianlong

Skype: brian.long.co

LinkedIn: <http://www.linkedin.com/in/brianlong>

#BOCC

Common Problems

- 123 Main St S, Schenectady, NY 12345
- 123 S Main Street, Schenectady, NY
- 123 Main St South, Schenectady, NY 12345
- 123 South Main Street, Schenectady, NY 12345
- 123 Main St S, Schenectady, NY 12345
- 123 S Main St, Skanecktidy, NY 12346

Cliff Clavin



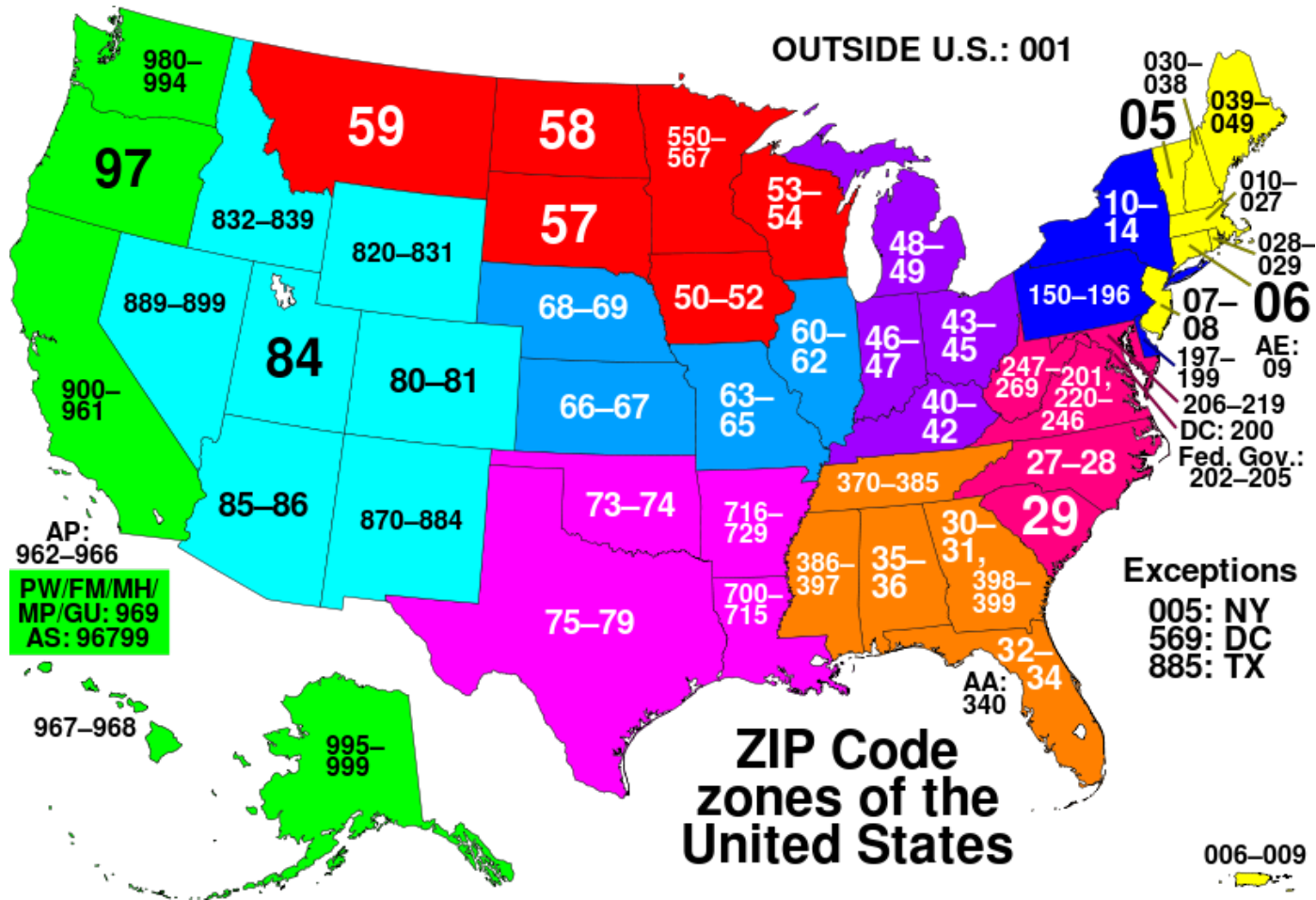
- http://en.wikipedia.org/wiki/Cliff_Clavin

Postal Barcode



ZIP Code
+4 Code
Delivery Point Code

Zip Code Zones



80301



ZIP+4



http://en.wikipedia.org/wiki/Terraced_house

Highrise +4



<http://en.wikipedia.org/wiki/Condominium>

Delivery Point



http://en.wikipedia.org/wiki/Single_family_house

Final Delivery

Wikimedia Foundation Inc.
200 2nd Ave. South #358
St. Petersburg, FL 33701-4313



33701-4313-00 UNIT-358

Unique Key



33701431300UNIT358



Data Hygiene

Use 3rd Party Software to cleanup addresses:

1 Microwsoft Weigh, Redmund, WA =>

1 Microsoft Way, Redmond, WA 98052-8300
[98052830099]

- Coding Accuracy Support Software (CASS)
- A list of CASS certified software vendors can be obtained from the USPS or by searching online.

Summary

- Street addresses are messy data.
- The post office provides a unique key that you can use for analysis.
- Add a data hygiene step using third party CASS software.
- Include the standardized data & bar code elements in your dataset.
- Create an indexed key using the bar code data.
- Search, sort, and group with ease.

Questions?

- What if a property doesn't have a street address (e.g. vacant land)?
- What if the address is incomplete?
- Are there edge cases to be aware of?

Presentation & notes at https://github.com/brianlong/street_address

P.S. I will offer free CASS service for open data projects.