

Evaluating the reliability of ACS data for transportation planning

Addison Larson

Delaware Valley Regional Planning Commission
ACS Data Users Conference

2019-05-15

Starting point

Analysis of every residence-based table without cross-tabulations in the latest CTPP release for five geographies

A101100

Total population (All persons)

Tract

Descriptive statistics

Minimum	Median	Mean	Maximum
0	3965	4134	11885

Minimum	Median	Mean	Maximum
0.13	4.45	5.65	151.98

Table 3: CVs by Reliability Bin

CV	Count
0-15	1361
15.1-30	6
30.1-60	2
60.1+	10

Table 4: High CVs

High CV Flag	Count
No	1369
Yes	10

Spatial distribution of CVs

Coefficients of Variation



High CVs



A102108

9:30 a.m. to 9:44 a.m.

TAZ

Descriptive statistics

Minimum	Median	Mean	Maximum
0	0	11	150

Minimum	Median	Mean	Maximum
15.2	100	95.31	319.15

Table 3: CVs by Reliability Bin

CV	Count
15.1-30	4
30.1-60	274
60.1+	2862

Table 4: High CVs

High CV Flag	Count
No	278
Yes	2862

Spatial distribution of CVs

Coefficients of Variation



High CVs

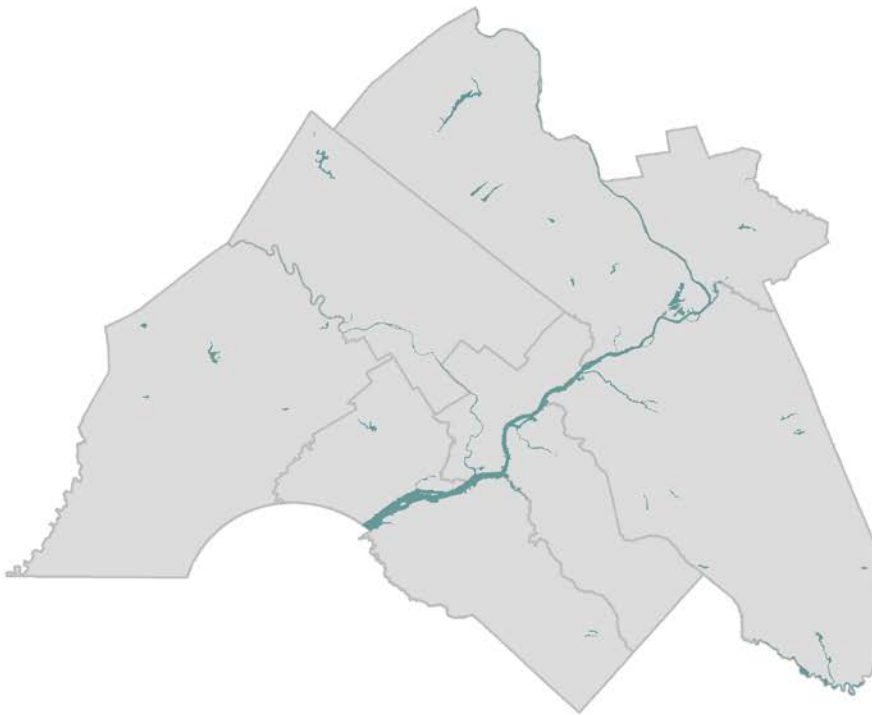


Outline

- Introduction to the CTPP
- Margins of error and why they matter
- CTPP coefficients of variation (CVs) by:
 - Table type
 - Geography
 - Variable detail
- A caveat on local context
- Recommendations, resources, references

Study area

DVRPC Region



Philadelphia



Census Transportation Planning Products

- Special tabulation of the ACS for transportation and planning purposes
- Unique table types, geographies, and variables
- Data based on 2012-2016 ACS 5-Year Estimates released 2019-04-02

CTPP table types

Census Tract 4.02

- Residence-based
- Workplace-based
- Flows

Example: No. of workers, Census Tract 4.02



CTPP geographies

County



POWPUMA



PUMA



TAD



Tract



TAZ



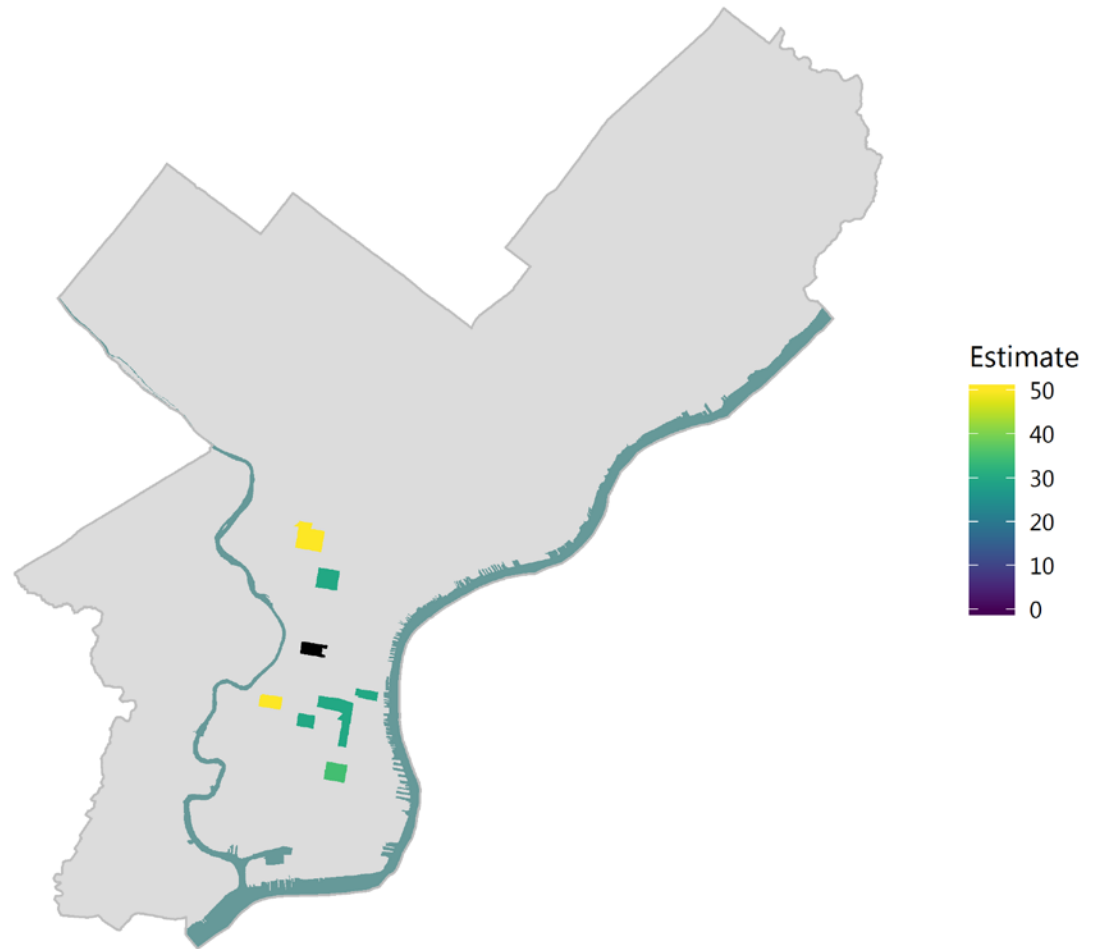
CTPP variables

Estimated bicycle commuters by census tract
Black tract is destination

At right:
Commuting flows
by means of
transportation

Other examples:

- Travel time by means of transportation
- Time arriving at work



MOEs make a difference

Using the previous slide's example:

Origin	Estimate	MOE	CV
Tract 20	50	65	79%
Tract 152	50	73	89%
Tract 41.02	35	58	101%
Tract 140	30	37	75%
Tract 28.01	30	27	55%
Tract 30.02	30	37	75%
Tract 16	30	37	75%
Tract 24	30	38	77%

CVs and the CTPP

- What's an "acceptable" CV? Rules of thumb:
 - 10-12% (Citro & Kalton, 2007)
 - Up to 15% (Francis et al., 2012)
- Choices in table type, geography, and variable detail simultaneously affect CVs
- Most of these choices apply to ACS data as well

CTPP CVs by table type

- Use Flows tables with caution

CV	Residence	Workplace	Flows
0-15%	366 (81%)	380 (81%)	1,041 (30%)
15.1-30%	51 (11%)	43 (9%)	525 (15%)
30.1-60%	23 (5%)	31 (7%)	595 (17%)
60.1+%	10 (2%)	14 (3%)	1,322 (38%)
Total obs.	450	468	3,483

CTPP CVs by geography

- Estimates at small geographies are often less reliable
- Selecting tract or TAZ carries reliability penalties

CV	County	PUMA	TAD	Tract	TAZ
Min	0.31%	0.75%	0.85%	3.09%	3.34%
Med	0.42%	1.01%	1.51%	6.93%	14.1%
Mean	0.45%	1.19%	1.63%	8.68%	23.62%
Max	0.64%	2.29%	3.37%	182.37%	218.84%

CTPP CVs by variable detail

- Use cross-tabulations with caution
- Below: zero-car households v. zero-car households × no. of workers in household

CV	ZC HH	ZC HH with 1 worker
Min	6.73%	11.85%
Med	16.59%	33.25%
Mean	22.06%	41.86%
Max	151.98%	319.15%

CTPP CVs by variable detail (cont.)

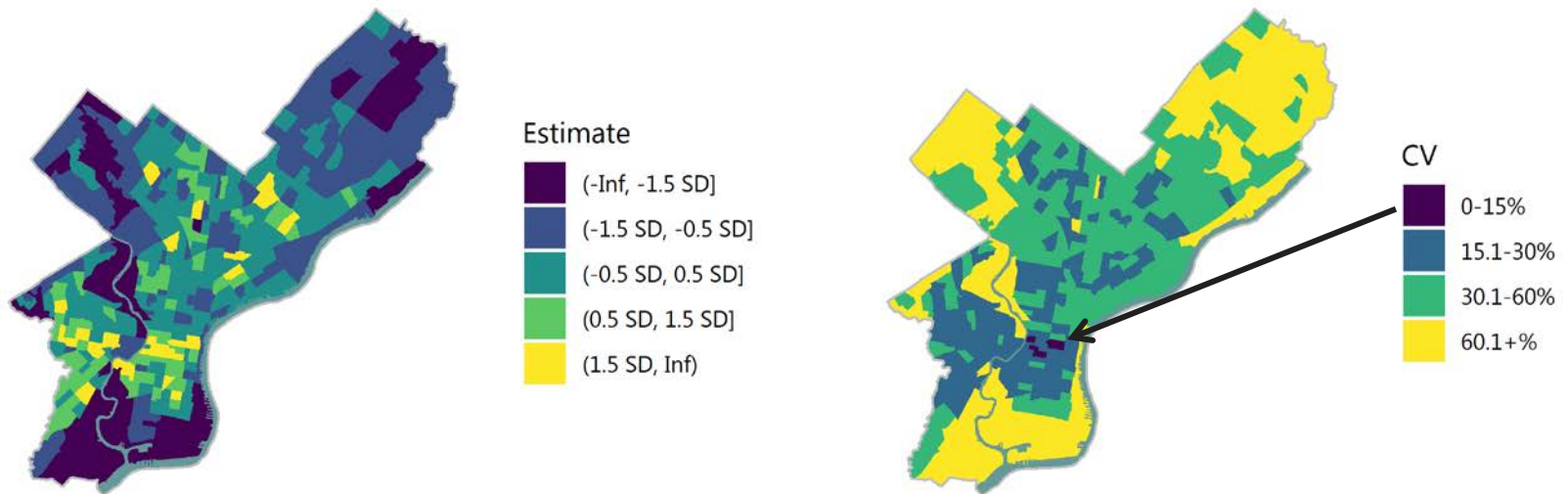
Using the previous slide's example:

	Estimate, zero-car households				
CV	$(-\infty, -1.5SD]$	$(-1.5SD, -0.5SD]$	$(-0.5SD, 0.5SD]$	$(0.5SD, 1.5SD]$	$(1.5SD, \infty)$
0-15%	0	5	56	61	30
15.1-30%	0	79	86	14	2
30.1-60%	0	34	0	0	0
60.1+%	14	3	0	0	0
	Estimate, zero-car households with one worker				
0-15%	0	0	0	1	4
15.1-30%	0	4	60	55	27
30.1-60%	0	72	99	6	0
60.1+%	18	36	2	0	0

A caveat on context

- Datasets with low reliability overall *might* be adequately reliable in your particular study area or for your particular research question

Zero-car households with one worker



Recommendations

- Careful selection of table type, geography, and variable
- Collapse geographies and/or subgroups
 - *Understanding and using ACS data (2018)*
- Cartographic choices
 - Francis et al. (2012)
 - *Map reliability calculator*
- Custom aggregate geographies
 - NYC Neighborhood Tabulation Areas
- Data-driven regionalization
 - Spielman & Folch (2015)

Resources

- Paper: *Evaluating the reliability of ACS data for transportation planning* [[link](#)]
- Online Appendix*: *Summary of data reliability for residence-based tables in the 2016 CTPP release* [[link](#)]
- Interactive map reliability calculator [[link](#)] →
- Email me: alarson@dvrpc.org

Map Classification Error Calculator

Questions? Click [here](#) for a how-to.

Upload two-column CSV file, where first column is estimate and second is MOE:

Browse... A112109_5_tract.csv
Upload complete

- File has a header
 Include estimates of 0 in error calculations

Select number of classes:
7

Error threshold (Default is 10%):
10

Custom comma-delimited breaks (Optional):
500,1000,1500,2000,2500,3000,3500

Here's the data you uploaded:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0	1060	1460	1524	1930	4035

User-Defined Break Values

500 1000 1500 2000 2500 3000 3500

Equal Interval Break Values

576.429 1152.857 1729.286 2305.714 2882.143 3458.571

Jenks Break Values

525 990 1340 1705 2110 2600

Quantile Break Values

Summary of Expected E

Equal Interval Breaks

Classes	Overall Error
2	OK: 2.22%
3	OK: 3.89%
4	OK: 4.87%
5	OK: 5.21%
6	OK: 6.76%
7	OK: 7.81%

Standard Deviation Breaks

Classes	Overall Error
2	NA
3	OK: 4.78%
4	NA
5	OK: 6.61%
6	NA
7	OK: 6.89%

Error by Number of Clas

Select a number of classes to see details

Equal Interval Breaks

Overall Expected Classification Error
OK

7.805

Expected Error By Class

OK

	Lower Bound	Error	Upper Bo
Class 1		0.000	
Class 2		1.104	

References

1. American Association of State Highway and Transportation Officials. (2019). *2012-2016 5-Year CTPP*. Retrieved 2019-04-12, from <https://ctpp.transportation.org/2012-2016-5-year-ctpp>
2. Citro, C. F., & Kalton, G. (Eds.). (2007). *Using the American Community Survey: Benefits and challenges*. Washington, DC: National Academies Press.
3. Francis, J., Vink, J., Tontisirin, N., Anantsuksomsri, S., & Zhong, V. (2012). *Alternative strategies for mapping ACS Estimates and error of estimation*. Retrieved 2019-04-18, from <https://pad.human.cornell.edu/papers/index.cfm>
4. New York City Department of City Planning. (n.d.). *Map reliability calculator*. Retrieved 2019-04-18, from <https://www1.nyc.gov/site/planning/data-maps/nyc-population/geographic-reference.page>
5. Spielman, S. E., & Folch, D. C. (2015). Reducing uncertainty in the American Community Survey through data-driven regionalization. *PLOS One*, *10*(2), e0115626.
6. United States Census Bureau. (2018). *Understanding and using American Community Survey data: What all data users need to know*. Retrieved 2019-04-20, from <https://www.census.gov/programs-surveys/acs/guidance/handbooks/general.html>

Source tables for CVs by table type

Description	Residence	Workplace	Flows
Total workers	A102101	A202100	A302100
Age of worker	A102102	A202101	B302101
Industry	A102105	A202104	B302102
Means of transportation	A102106	A202105	A302103
Travel time	A102110	A202113	B302106

Universe: Workers 16 years and over