Small Area Health Indicators with ACS 5-Year Estimates

ACS Data Users Conference
May 12, 2015

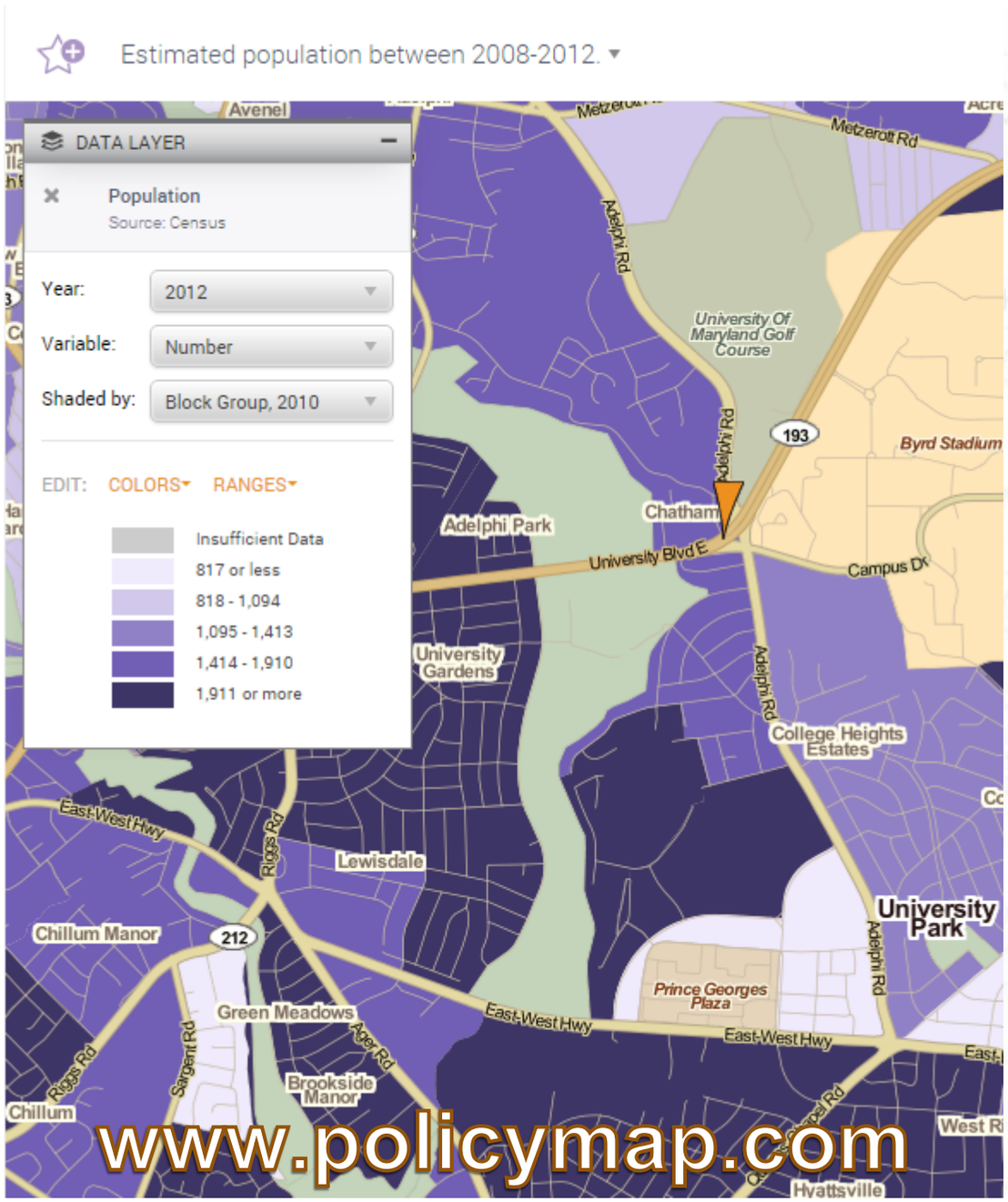
Morgan Robinson
Data Analyst, PolicyMap

@policymap **#ACSConf15**

TRF policymap

Agenda

- Introductions
- Local Health Indicators
- Behavioral Risk Factor Surveillance System
- Creating Small Area Estimates
- Q & A


Estimated population between 2008-2012.
DATA LAYER
Population
Source: Census
Year: 2012
Variable: Number
Shaded by: Block Group, 2010
EDIT: COLORS RANGES
Insufficient Data
817 or less
818 - 1,094
1,095 - 1,413
1,414 - 1,910
1,911 or more
Avenel
Metzerott Rd
Adelphi Rd
University Of Maryland Golf Course
193
Byrd Stadium
Chatham
Adelphi Park
University Blvd E
Campus Dr
University Gardens
College Heights Estates
East-West Hwy
Riggs Rd
Lewisdale
Chillum Manor
212
University Park
Prince Georges Plaza
Green Meadows
Sargent Rd
Ager Rd
Brookside Manor
Chillum
Hyattsville
www.policymap.com


1. GROWING NEED, GROWING USAGE
8,000,000
maps were made on PolicyMap since 2011
2. FREE SERVICES FOR THE PUBLIC
FREE
79%
of our user base uses PolicyMap for free
3. DATA SAVVY FUTURE LEADERS
1,082,663
students are equipped with university access
4. DATA-DRIVEN GOVERNMENT
2,469
government agencies use PolicyMap
5. SMARTER NONPROFITS
40%
of the organizations using PolicyMap are nonprofits or philanthropies
6. FINANCIALLY SUSTAINABLE
64%
of our revenue is product generated

Health Data

- Access, Quality of Care
- Health Services
- Intersections, Social Determinants of Health
- Health Outcomes
- Risk Factors and Behaviors

Health Data: Limitations

- Availability
- Privacy, Suppression
- Comparability: across time and place

Good Conditions for Creating Estimates

- Robust survey
- Documented link between demographics and variables
- Demographic/socioeconomic classification
- Geographic information provided
- Clearly-coded variables

CDC Behavioral Risk Factor Surveillance System

- ✓ Robust survey
- ✓ Documented link between demographics and variables
- ✓ Demographic/socioeconomic classification
- ✓ Geographic information provided
- ✓ Clearly-coded variables

About how long has it been since you last visited a doctor for a routine checkup?

Have you EVER been told by a doctor, nurse or other health professional that you have high blood pressure?

CDC Behavioral Risk Factor Surveillance System

During the past 30 days, how many days per week or per month did you have at least one drink of any alcoholic beverage such as beer, wine, a malt beverage or liquor?

- Annual phone survey of 400,000
- Record-level data available
- 53 States (including DC, Guam, PR), some cities/counties available
- Topic modules by state

During the past month, how many times per day, week, or month did you eat dark green vegetables?

Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?

CDC Behavioral Risk Factor Surveillance System

- http://www.cdc.gov/brfss/annual_data/annual_2013.html
- Text (84 MB) or SAS (124 MB) files available
- Combined Landline and Cellphone surveys
- Survey weights included
  - Create weighted estimates
  - Sampling

Creating Small Area Estimates

- Predict a Yes/No response to survey questions
- Use variables in BRFSS –*and*– ACS
- Consider distribution of responses

Multilevel, Mixed-effects Regression and Post-stratification

- Model individual survey responses as a function of demographic and geographic predictors
- Estimate the probability of Yes for “cells” of individual-level demographic variables
- Good tutorials:
  Jonathan Kastellec: http://www.princeton.edu/~jkastell/mrp_primer.html;
  Jared Knowles: http://jaredknowles.com/journal/2013/11/25/getting-started-with-mixed-effect-models-in-r;
  Bodo Winter: http://www.bodowinter.com/tutorials.html

ACS 5-Year Estimates for Post-stratification

- Population counts for individual-level demographic variables
- Age of householder by household income by race
- Estimate the probability of Yes for “cells” of individual-level demographic variables
- Other options:
  - Sex by age by race
  - Age by disability by poverty status

ACS 5-Year Estimates for Post-stratification

- Population counts for individual-level demographic variables
- Age of householder by household income by race
- Desired result: cross-tab:

For prediction:

	FIPS	MSLABEL	AGE_GLABEL	RACELABEL	INCOMELABEL	STATE
1	34	SC	A3	RBK	I3	NJ
2	24	SC	A3	RBK	I3	MD
3	42	SC	A3	RBK	I3	PA
4	10	SC	A3	RBK	I3	DE
5	34	SC	A2	RMT	I3	NJ
6	24	SC	A2	RMT	I3	MD

For post-stratification:

FIPS	MSCODE	RACE	AGE	INCOME	SUBPOP	PCT_POP
34019	2	RWH	3	8	16,972	0.362526
48269	4	RWH	3	8	43	0.34127
08039	3	RWH	3	8	2,748	0.333495
48301	4	RWH	3	8	10	0.30303
24009	3	RWH	3	8	9,360	0.302687

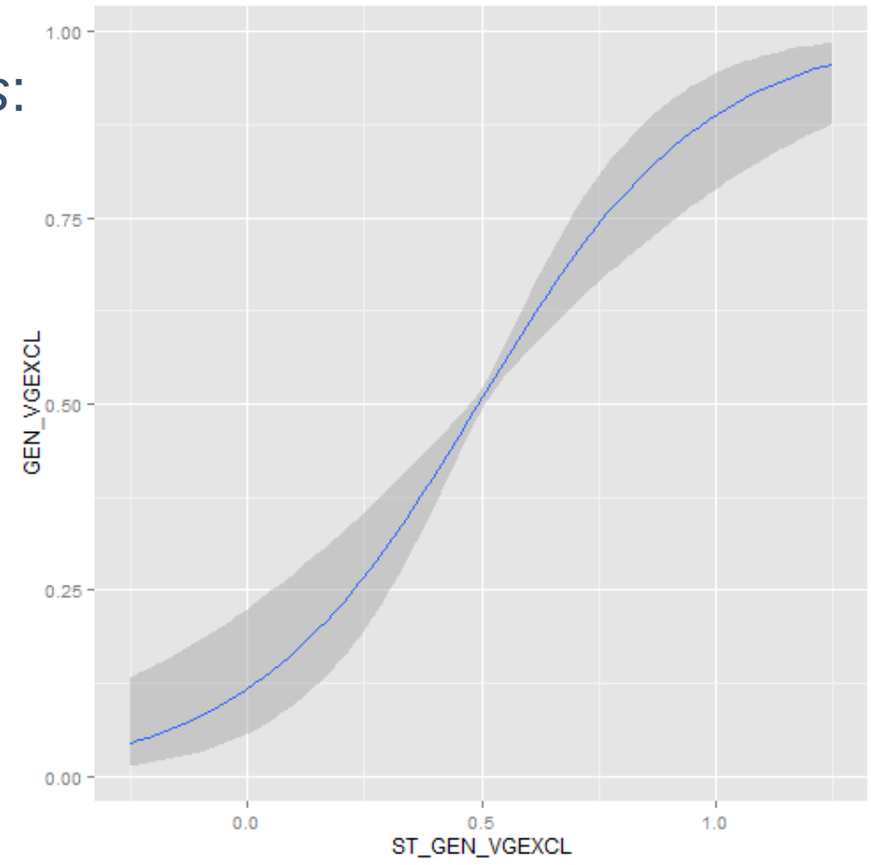
ACSSF	B19037I	63	69 CELLS	AGE OF HOUSEHOLDER BY HOUSEHOLD INCOME IN THE PAST 12 MONTHS (IN 2013 INFLATION-ADJUSTED DOLLARS) (HISPANIC OR LATINO HOUSEHOLDER)
ACSSF	B19037I	63		Universe: Households with a householder who is Hispanic or Latino
ACSSF	B19037I	63	1	Total:
ACSSF	B19037I	63	2	Householder under 25 years:
ACSSF	B19037I	63	3	Less than $10,000
ACSSF	B19037I	63	4	$10,000 to $14,999
ACSSF	B19037I	63	5	$15,000 to $19,999
ACSSF	B19037I	63	6	$20,000 to $24,999
ACSSF	B19037I	63	7	$25,000 to $29,999
ACSSF	B19037I	63	8	$30,000 to $34,999
ACSSF	B19037I	63	9	$35,000 to $39,999
ACSSF	B19037I	63	10	$40,000 to $44,999
ACSSF	B19037I	63	11	$45,000 to $49,999
ACSSF	B19037I	63	12	$50,000 to $59,999
ACSSF	B19037I	63	13	$60,000 to $74,999
ACSSF	B19037I	63	14	$75,000 to $99,999
ACSSF	B19037I	63	15	$100,000 to $124,999
ACSSF	B19037I	63	16	$125,000 to $149,999
ACSSF	B19037I	63	17	$150,000 to $199,999
ACSSF	B19037I	63	18	$200,000 or more

FIPS	County	Principal City	County Status	MSCODE
06037	Los Angeles	55%	Central	CC
08039	Elbert	0%	Outlying	SC
17031	Cook	56%	Central	CC
42101	Philadelphia	100%	Central	IC
48269	King	0%	NULL	RA

Modeling the Survey Data

For a sample R project, visit http://git.io/vJn0v

- *Would you say that in general your health is*:
  - ***Excellent***
  - ***Very Good***
- GEN_VGEXCL as a function of:
  - Statewide percentage
  - Income
  - Age
  - Race
  - Metro classification



Modeling the Survey Data

For a sample R project, visit http://git.io/vJn0v

- Data sample (using weights)
- Table of statewide values for variable
- A second sample, for validating the model
- Shell of age (4) x income (8) x race (5) x metro (4)
  - 640 cells per state

Modeling the Survey Data

```
glm <- glmer(GEN_VGEXCL ~  RACELABEL + INCOMELABEL + AGE_GLABEL + ST_GEN_VGEXCL + (1| MSLABEL:RACELABEL),
family = "binomial", data = data)

predictsample <- data[sample(1:nrow(data), size=5000, replace=T, prob = as.integer(data$LLCPWT)), ]
predictsample$PR_GEN_VGEXCL <- predict(
  glm, type = "response", newdata = predictsample, allow.new.levels = TRUE)
summary(predictsample$PR_GEN_VGEXCL)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1245  0.3746  0.5320  0.5027  0.6241  0.7984
```

```
data <- data[complete.cases(data), ]
head(data)
```

```
##   FIPS STATE MSLABEL INCOMELABEL     LLCPWT AGE_GLABEL RACELABEL
## 1   12    FL       CC          I6   755.2860         A4        RWH
## 2   39    OH       IC          I6  2109.2465         A2        RWH
## 3    6    CA       IC          I3 11698.2597         A2        RHS
## 4   55    WI       RA          I8  1016.0006         A4        ROT
## 5    6    CA       IC          I8  1871.4452         A3        RWH
## 6   36    NY       CC          I8   555.7514         A4        RWH
##   GEN_VGEXCL ST_GEN_VGEXCL
## 1          0     0.4957354
## 2          0     0.5022108
## 3          0     0.5072307
## 4          0     0.5369327
## 5          0     0.5072307
## 6          1     0.5159127
```

Modeling the Survey Data

For a sample R project, visit http://git.io/vJn0v

- Predict a response for each cell (640 per state)

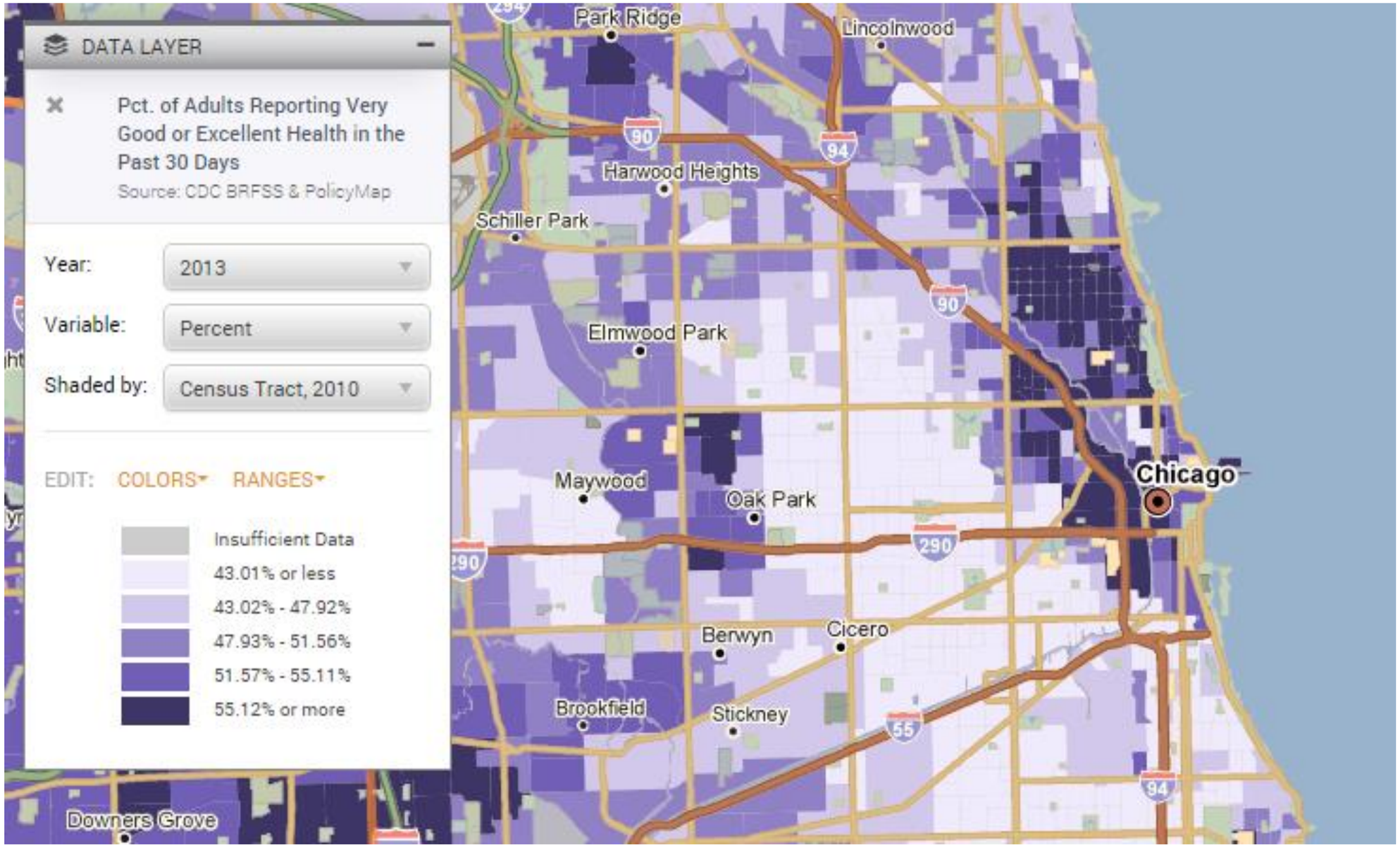
```
##   FIPS MSLABEL AGE_GLABEL RACELABEL INCOMELABEL STATE ST_GEN_VGEXCL
## 1   34      CC         A3        RBK          I3    NJ     0.5280175
## 2   24      CC         A3        RBK          I3    MD     0.5368445
## 3   42      CC         A3        RBK          I3    PA     0.4265109
## 4   10      CC         A3        RBK          I3    DE     0.4444476
## 5   34      CC         A2        RMT          I3    NJ     0.5280175
## 6   24      CC         A2        RMT          I3    MD     0.5368445
##   PR_GEN_VGEXCL
## 1     0.2970425
## 2     0.3027684
## 3     0.2359465
## 4     0.2460825
## 5     0.2822745
## 6     0.2878322
```

ACS 5-Year Estimates for Post-stratification

Tract	POP > 18	%	#
010100	3719	45.11	1677
010201	5325	44.79	2385

- Translate individual predictions into place values
- Calculate weighted average probability for each geography
- Calculate estimates (number and percent) for adults 18+

FIPS	MSCODE	RACE	AGE	INCOME	SUBPOP	PR_GEN_VGEXCL
06037	2	RBK	1	2	1282	0.3843
06037	2	RBK	1	6	1369	0.6019
06037	2	RBK	1	5	1764	0.5345
17031	2	RHS	1	4	1208	0.4038
06037	2	RWH	1	4	2199	0.5762
17031	2	RHS	1	3	1317	0.3687
06037	2	RWH	1	3	2205	0.5396
17031	2	RBK	1	5	1403	0.5284


DATA LAYER
Pct. of Adults Reporting Very Good or Excellent Health in the Past 30 Days
Source: CDC BRFSS & PolicyMap
Year: 2013
Variable: Percent
Shaded by: Census Tract, 2010
EDIT: COLORS RANGES
Insufficient Data
43.01% or less
43.02% - 47.92%
47.93% - 51.56%
51.57% - 55.11%
55.12% or more
Park Ridge
Lincolnwood
Harwood Heights
Schiller Park
Elmwood Park
Maywood
Oak Park
Chicago
Berwyn
Cicero
Brookfield
Stickney
Downers Grove
90
94
290
55

For more information:

- www.policymap.com/
- For a sample R project, visit http://git.io/vJn0v
- http://www.cdc.gov/brfss/annual_data/annual_2013.html

Thank you!

Small Area Health Indicators
with ACS 5-Year Estimates:
a presentation to the ACS Data Users Conference
May 12, 2015

Morgan Robinson
Data Analyst, PolicyMap

@policymap **#ACSConf15**