

# Applications of R Shiny to Evaluate and Improve Total Survey Quality

Xiaodan Lyu

Joint work with Heike Hofmann, Emily Berg

# Overview

1. Introduction
2. National Resource Inventory (NRI)
3. *iNtr: an interactive NRI table review tool*
4. *viscover: visualize soil and cropland data layer*
5. Discussion

# Introduction

- Total survey quality
  - \* the concept of “Total Survey Error”. (Groves and Lyberg, 2010)
  - \* managing data quality in a National Statistical Office.  
(Brackstone, 1999)

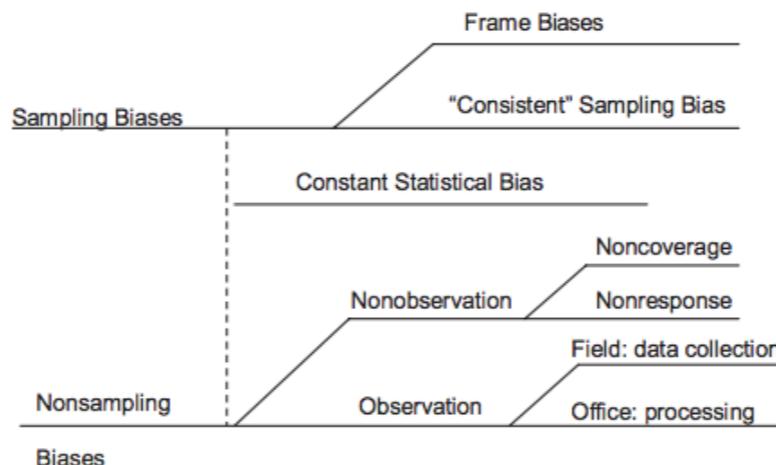


Figure 2. Schematic Presentation in Kish (1965) of Biases in Surveys.

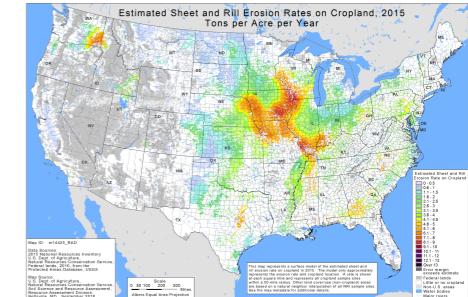
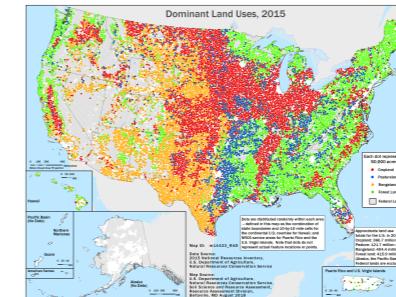
# Introduction

- Graphical displays aids the assessment
  - \* EDA & data quality of administrative data (Karr et al., 2006)
  - \* heatmap & weighting procedures in calibration (Williams and Berg, 2013)
- Application of R Shiny to survey
  - \* conduct a social-network based survey (Joblin and Mauerer, 2016)
  - \* examples:
    - evaluate output of an estimation procedure (NRI)
    - review input to a model-assisted estimation procedure (SAE)

iNtr: an **interactive NRI**  
**table review tool**

# National Resource Inventory

- A longitudinal survey
  - \* monitor national resources on non-federal US land
  - \* collect data based on aerial photographs of sampled segments
- National and state level estimates
  - \* surface area by land cover/use
  - \* average water and wind erosion on cropland and pastureland
- Record level data set (Pointgen)
  - \* location with a single weight and complete data



**Table 2 - Land Cover/use of non-Federal rural land, by State and year**  
**In thousands of acres, with margins of error**

<b>State</b>	<b>Year</b>	<b>Cropland</b>	<b>CRP land</b>	<b>Pastureland</b>	<b>Rangeland</b>	<b>Forest land</b>	<b>Other rural land</b>	<b>Total rural land</b>
<b>Alabama</b>	<b>1982</b>	4,464.7 ±176.9	—	3,793.9 ±186.4	53.7 ±47.3	20,876.8 ±184.4	523.5 ±73.2	29,712.6 ±112.8
	<b>1987</b>	3,944.7 ±187.9	207.5 —	3,643.8 ±157.9	52.8 ±45.6	21,160.9 ±181.1	491.2 ±74.0	29,500.9 ±116.9
	<b>1992</b>	3,126.2 ±192.4	535.2 —	3,753.3 ±147.5	52.7 ±45.6	21,250.7 ±189.7	611.7 ±84.8	29,329.8 ±124.5
	<b>1997</b>	2,915.5 ±209.9	522.2 —	3,558.0 ±134.4	53.8 ±46.7	21,325.9 ±200.6	590.9 ±78.9	28,966.3 ±139.8
	<b>2002</b>	2,508.5 ±183.0	504.6 —	3,452.2 ±194.7	50.7 ±108.9	21,550.4 ±250.2	505.8 ±83.3	28,572.2 ±151.7
	<b>2007</b>	2,200.2 ±180.6	459.8 —	3,434.9 ±177.3	50.7 ±108.9	21,668.5 ±262.4	554.6 ±90.2	28,368.7 ±164.9
	<b>2012</b>	2,217.0 ±189.0	329.2 —	3,302.9 ±175.9	50.7 ±108.9	21,787.8 ±264.1	591.6 ±85.8	28,279.2 ±168.1
	<b>2015</b>	2,274.9 ±194.2	225.6 —	3,220.2 ±187.2	50.7 ±108.9	21,887.4 ±263.4	591.6 ±85.2	28,250.4 ±168.1
<b>Arizona</b>	<b>1982</b>	1,253.0 ±146.6	—	83.6 ±50.4	33,366.4 ±1,026.6	4,572.6 ±862.1	1,711.8 ±577.2	40,987.4 ±270.6
	<b>1987</b>	1,234.6 ±145.5	0.0 —	76.6 ±40.2	33,395.7 ±1,046.3	4,553.8 ±860.8	1,784.4 ±594.7	41,045.1 ±286.7
	<b>1992</b>	1,199.9 ±148.8	0.0 —	83.3 ±36.0	33,796.0 ±1,069.5	4,434.9 ±881.6	1,801.6 ±569.8	41,315.7 ±295.4

**Table 14 - Estimated average annual sheet and rill erosion on non-Federal rural land,  
by State and year**  
**Tons per acre per year with margins of error**

State	Year	Cropland			CRP land	Pastureland
		Cultivated	Non-Cultivated	Total		
<b>Alabama</b>	<b>1982</b>	5.02 ±0.29	0.41 ±0.25	4.73 ±0.28	--	0.65 ±0.08
	<b>1987</b>	4.39 ±0.29	0.32 ±0.07	4.07 ±0.27	2.32 ±1.39	0.51 ±0.07
	<b>1992</b>	4.86 ±0.26	0.37 ±0.17	4.35 ±0.23	0.62 ±0.27	0.49 ±0.06
	<b>1997</b>	4.72 ±0.25	0.35 ±0.16	4.18 ±0.22	0.76 ±0.38	0.52 ±0.06
	<b>2002</b>	4.47 ±0.30	0.41 ±0.11	3.62 ±0.29	0.68 ±0.42	0.53 ±0.07
	<b>2007</b>	4.34 ±0.47	0.36 ±0.09	3.34 ±0.41	0.51 ±0.24	0.44 ±0.04
	<b>2012</b>	3.87 ±0.34	0.39 ±0.07	3.03 ±0.33	0.62 ±0.37	0.42 ±0.06
	<b>2015</b>	4.22 ±0.43	0.39 ±0.07	3.31 ±0.42	0.51 ±0.41	0.44 ±0.08
<b>Arizona</b>	<b>1982</b>	0.59 ±0.06	0.45 ±0.04	0.57 ±0.06	--	0.15 ±0.08
	<b>1987</b>	0.65 ±0.06	0.44 ±0.05	0.62 ±0.04	0.00 -	0.11 ±0.03
	<b>1992</b>	0.67 ±0.05	0.27 ±0.03	0.60 ±0.05	0.00 -	0.15 ±0.04

**“...all data collected prior to 2015 were simultaneously reviewed (edited) as 2015 NRI data were collected...”**

– Summary Report: 2015 National Resources Inventory

# 2015 NRI Table Review

- Reasons
  - \* Multiple estimation runs before final publication
- Differences
  - \* a 2015 estimation run VS final 2012 result
  - \* a new 2015 estimation run VS an earlier 2015 estimation
- Results
  - \* Expected changes: updated programming
  - \* Surprised changes: data issue or estimation issue

[https://lyux.shinyapps.io/table\\_review/](https://lyux.shinyapps.io/table_review/)

NRI Data Review Visualization

Info

+

Options

Please select V1 with new values:

Final\_2015

Please select V2 with old values:

Final\_2012

Table number

2

Table cell

level  se

Color scale

absolute relative difference  absolute difference

Table 2 (level): Land Cover/use of non-Federal rural land, by State and year, in thousands of acres.

Filter: diff > 0.1 of new val && new val >= 1

Difference Table - US

+

Difference Map - US

Apply filter

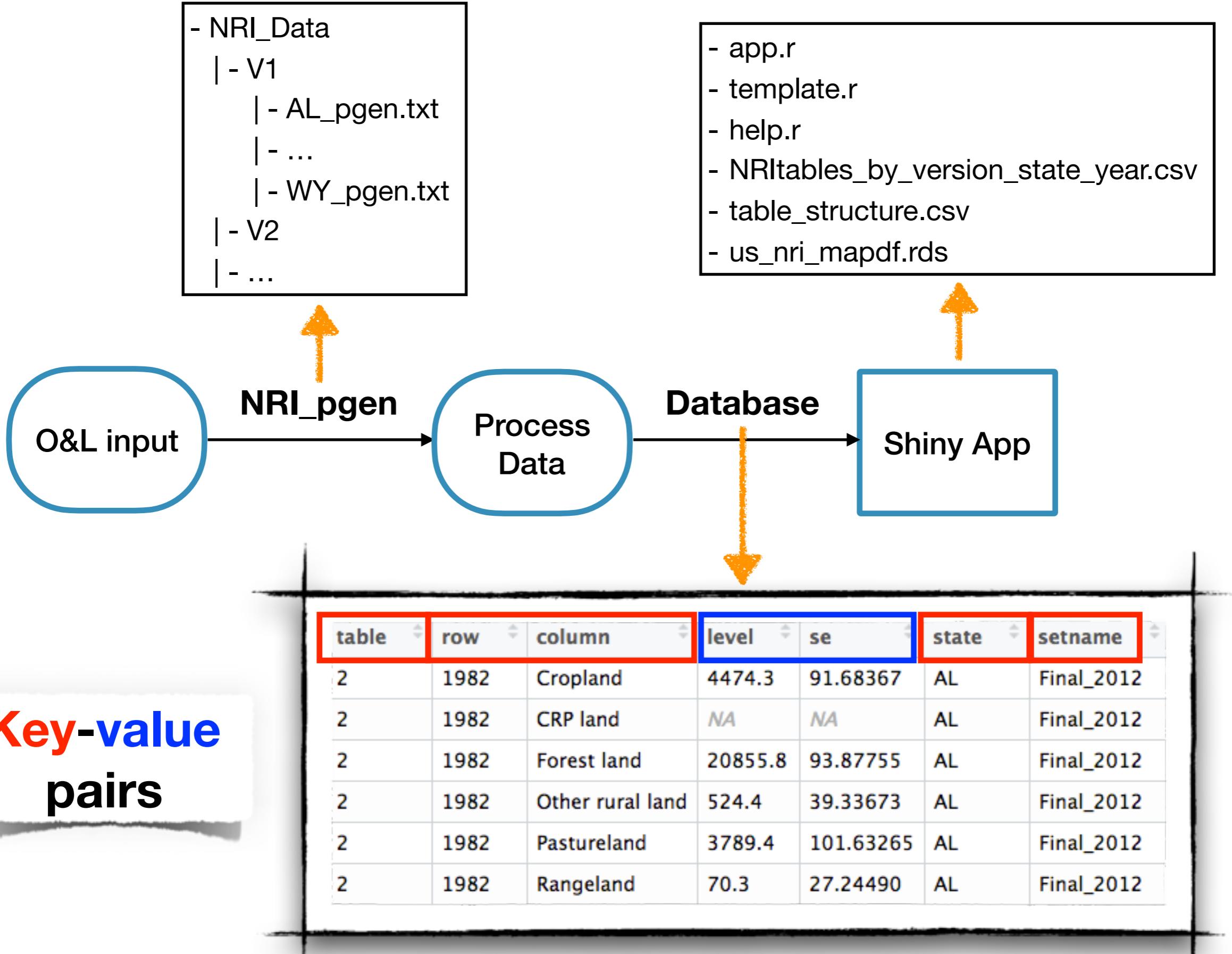
Please click a cell in the 1st panel (Difference Table - US) to see the corresponding difference map.

Difference Table - State

Apply filter

Enable hover

Please click a state polygon in the 2nd panel (Difference Table - State) to see the corresponding difference table.



# iNtr Skeleton

Info

Options

Panel 1  
Difference Table - US

Panel 2  
Difference Map - US

Panel 3  
Difference Table - State

## - NRItables\_by\_version\_state\_year.csv

table	row	column	level	se	state	setname
2	1982	Cropland	4474.3	91.68367	AL	Final_2012
2	1982	CRP land	NA	NA	AL	Final_2012
2	1982	Forest land	20855.8	93.87755	AL	Final_2012
2	1982	Other rural land	524.4	39.33673	AL	Final_2012
2	1982	Pastureland	3789.4	101.63265	AL	Final_2012
2	1982	Rangeland	70.3	27.24490	AL	Final_2012

### Options

Please select V1 with new values:

Final\_2015

Please select V2 with old values:

Final\_2012

Table number

2

Table cell

level  se

Color scale

absolute relative difference  absolute difference

Table 2 (level): Land Cover/use of non-Federal rural land, by State and year, in thousands of acres.

Filter: diff > 0.1 of new val && new val >= 1

## - table\_structure.csv

table	row	column	name	cond_level_rel	cond_level	cond_se_rel	cond_se	digit
2	Year	Broad/Cover Use	Land Cover/use of non-Federal rural land, by State and year, in thousands of acres.	> 0.1	$\geq 1$	$> 0.1$	$> 0$	1
3a	Year	Broad/Cover Use	Estimated average annual change in land cover/use of non-Federal rural land, by State and year, in thousands of acres.	$> 0.05$	$\geq 0.1$	$> 0.1$	$\geq 0.005$	3

## Options

Please select V1 with new values:

Final\_2015

Please select V2 with old values:

Final\_2012

Table number

2

Table cell

level  se

Color scale

absolute relative difference  absolute difference

Table 2 (level): Land Cover/use of non-Federal rural land, by State and year, in thousands of acres.

Filter: diff > 0.1 of new val && new val >= 1

table	row	column	lev
2	1982	Cropland	420588
2	1982	CRP land	
2	1982	Forest land	410289
2	1982	Other rural land	42776
2	1982	Pastureland	131289
2	1982	Rangeland	419356

```
> tb_dif %>% glimpse
```

Observations: 49

Variables: 16

\$ level.1

\$ level\_2

```
$ absdiff level
```

```
$ relabsdiff level
```

```
$ relabdiff.level fil
```

\$ adsu  
\$ sel sk

\$\rightarrow\$ relative

```
$ se.1
```

\$ se.2

\$ absdiff.se

\$ relabsdiff.se

```
$ absdiff.se.filter
```

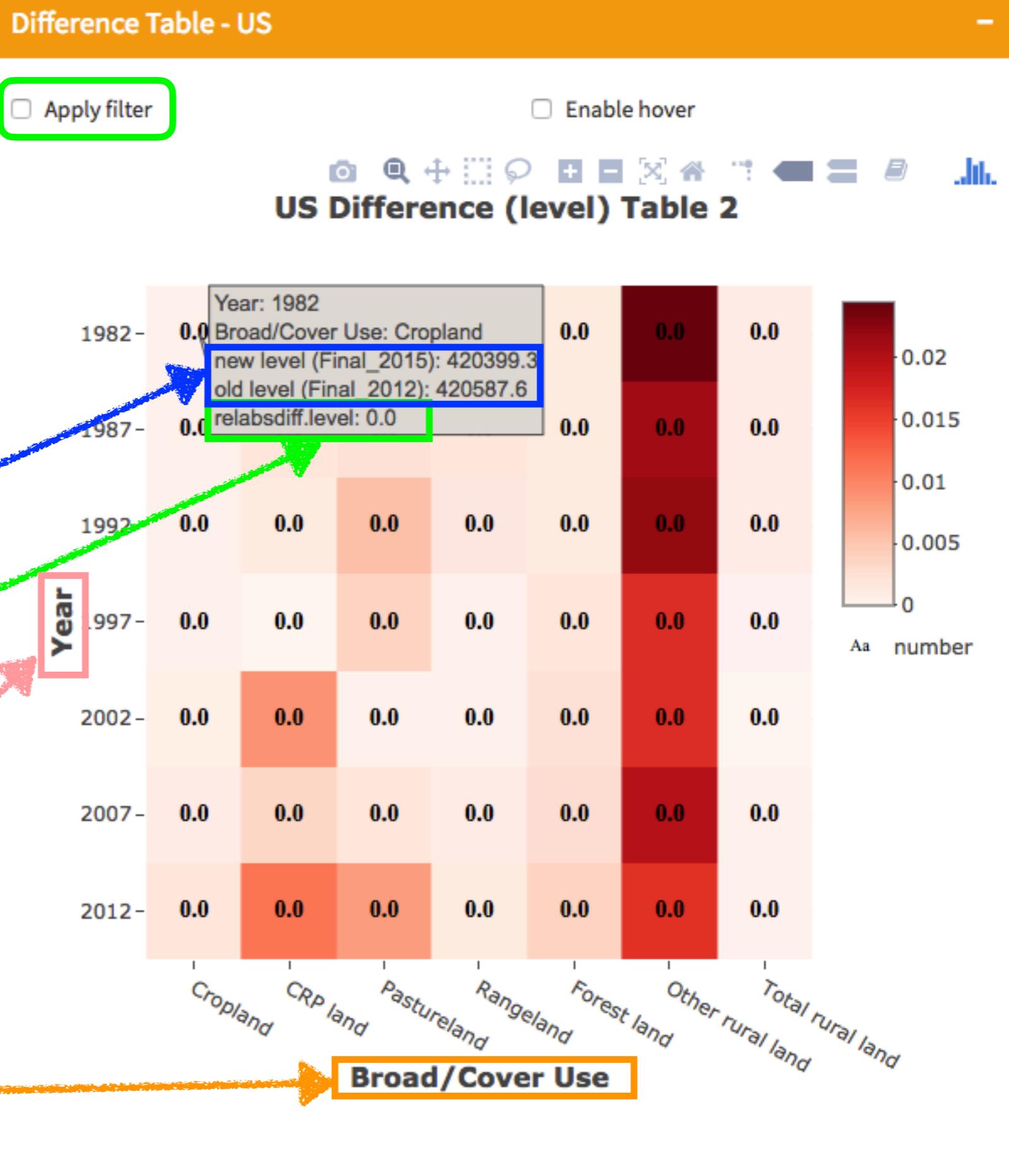
\$ relabsdiff.se.fil

```
$ table
```

\$ row

\$ column

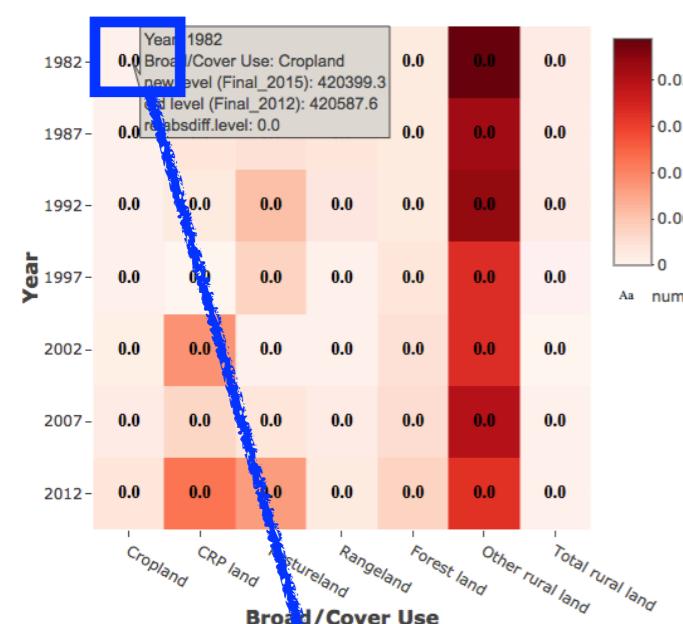
\$ state



## Difference Table - US

 Apply filter Enable hover

## US Difference (level) Table 2



Observations: 50

Variables: 16

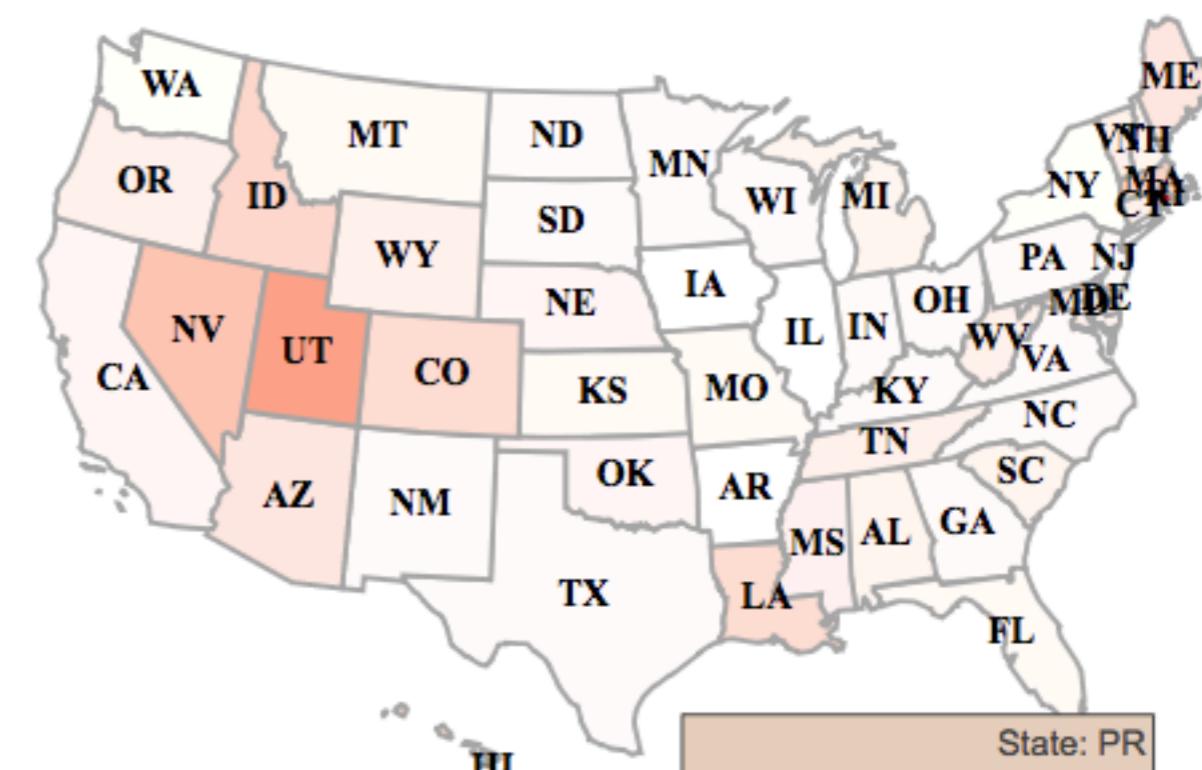
```
$ level.1 <dbl> 4464.7, 12
$ level.2 <dbl> 4474.3, 12
$ absdiff.level <dbl> 9.6, 7.7,
$ relabsdiff.level <dbl> 2.150200e-
$ absdiff.level.filter <dbl> NA, NA, 0.
$ relabsdiff.level.filter <dbl> NA, NA, 3.
$ se.1 <dbl> 90.25510,
$ se.2 <dbl> 91.68367,
$ absdiff.se <dbl> 1.4285714,
$ relabsdiff.se <dbl> 0.01582815
$ absdiff.se.filter <dbl> NA, NA, NA
$ relabsdiff.se.filter <dbl> NA, NA, NA
$ table <int> 2, 2, 2, 2
$ row <int> 1982, 1982
$ column <chr> "Cropland"
$ state <chr> "AL", "AZ"
```

## Difference Map - US

 Apply filter

## US Difference (level) Table 2

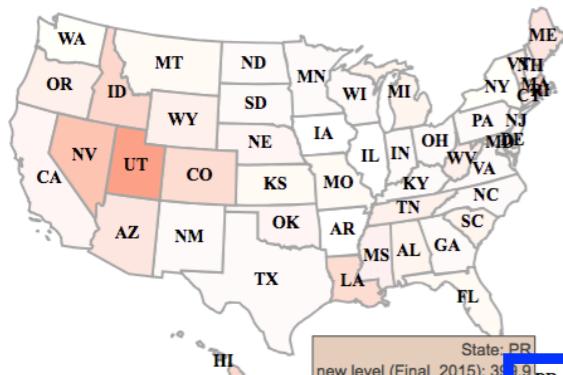
Year: 1982, Broad/Cover Use: Cropland



## Difference Map - US

 Apply filter

Year: 1982, Broad/Cover Use: Cropland



Observations: 49

Variables: 16

```
$ level.1 <dbl> 399.9, NA
$ level.2 <dbl> 402.0, NA
$ absdiff.level <dbl> 2.1, NA, 8
$ relabsdiff.level <dbl> 0.00525131
$ absdiff.level.filter <dbl> NA, NA, NA
$ relabsdiff.level.filter <dbl> NA, NA, NA
$ se.1 <dbl> 14.336735
$ se.2 <dbl> 15.357143
$ absdiff.se <dbl> 1.02040816
$ relabsdiff.se <dbl> 0.07117437
$ absdiff.se.filter <dbl> NA, NA, NA
$ relabsdiff.se.filter <dbl> NA, NA, NA
$ table <int> 2, 2, 2, 2
$ row <int> 1982, 1982
$ column <chr> "cropland"
$ state <chr> "PR", "PR"
```

## Difference Table - State

 Apply filter Enable hover

## PR Difference (level) Table 2

PR Difference (level) Table 2

Year	Cropland	CRP land	Pastureland	Rangeland	Forest land	Other rural land	Total rural land
1982-	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1987-	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1992-	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1997-	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2002-	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2007-	0.0	0.0	0.0	0.1	0.0	0.0	0.0
2012-	0.1	0.0	0.0	0.1	0.0	0.0	0.0

Aa number

Broad/Cover Use

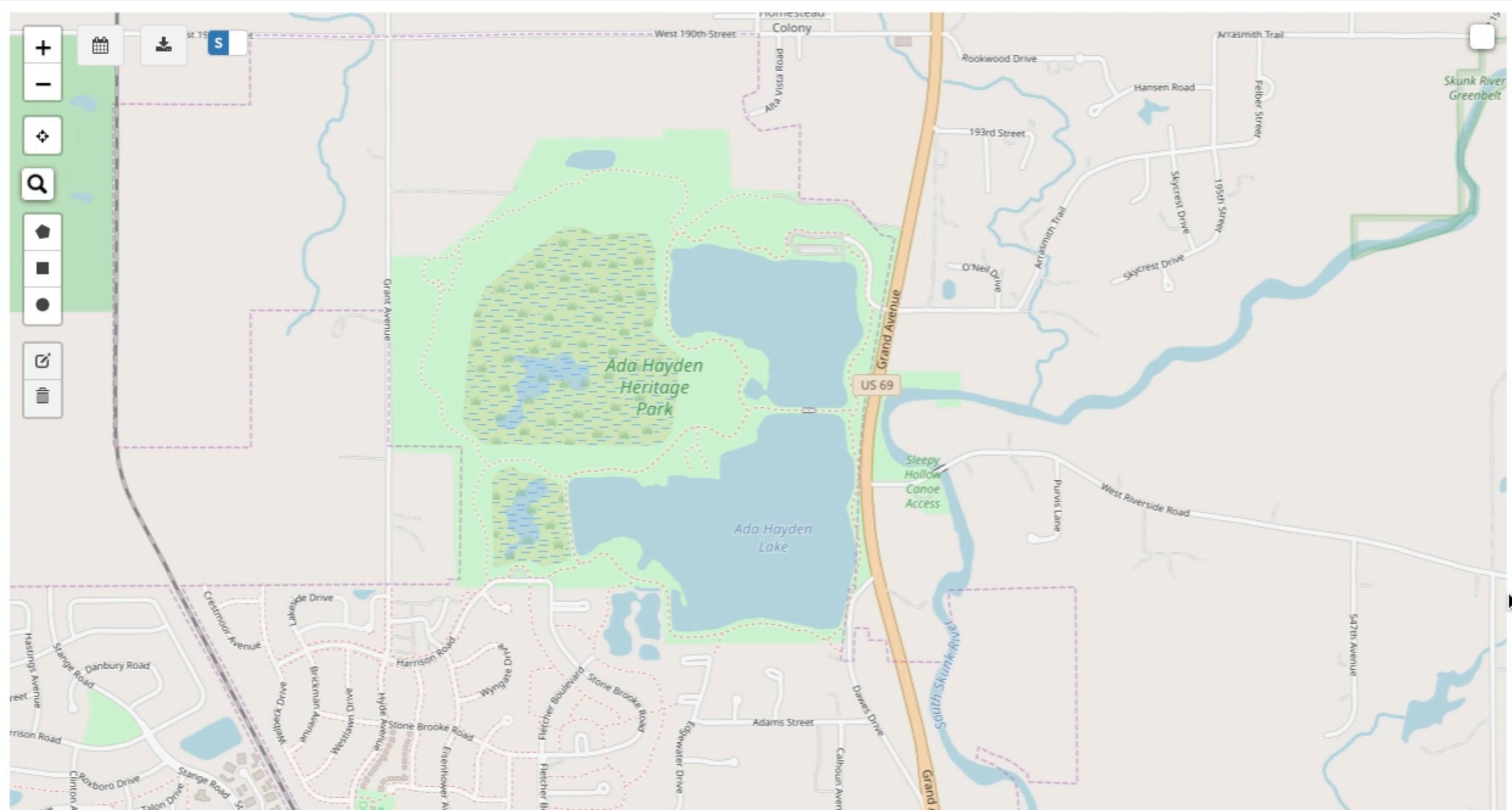
viscover: visualize soil  
and crop data and  
their overlay

# Conservation Effects Assessment Project (CEAP)

- CEAP sample consists of points sampled from NRI cropland points.
- An on-site study to quantify soil and nutrient loss from cropland and pastureland.
- Model based small area estimation
  - \* integrate population level soil and crop data to implement model (Lyu, Berg and Hofmann, submitted)
    - data quality of auxiliary variables
    - integrity of overlay operation

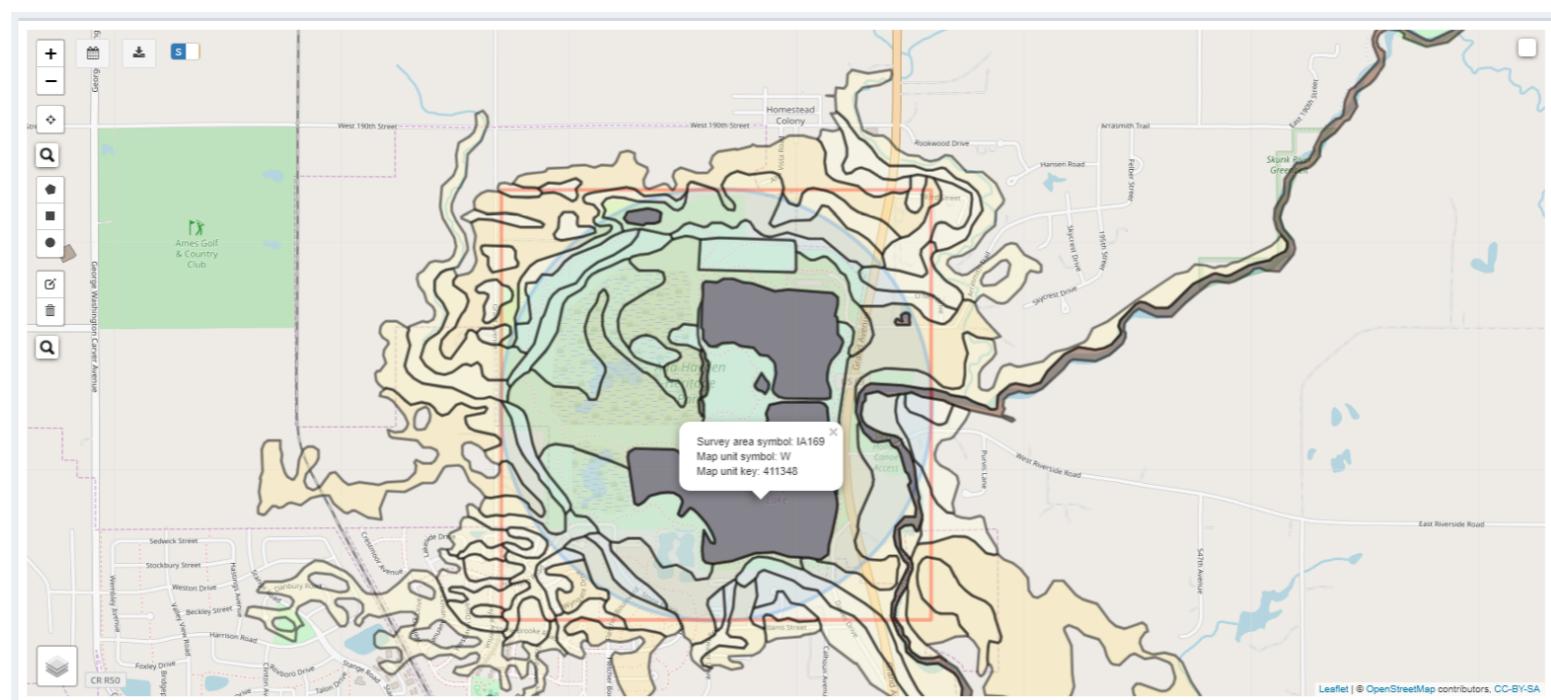
# Public databases

- Soil data layer (SDL)
  - \* maintained by USDA NRCS since 1896
  - \* tabular and spatial soil information available for the United States and the Territories
- Cropland data layer (CDL)
  - \* created by USDA NASS since 1997
  - \* created annually for the contiguous United States since 2008
  - \* geo-referenced crop-specific land cover data



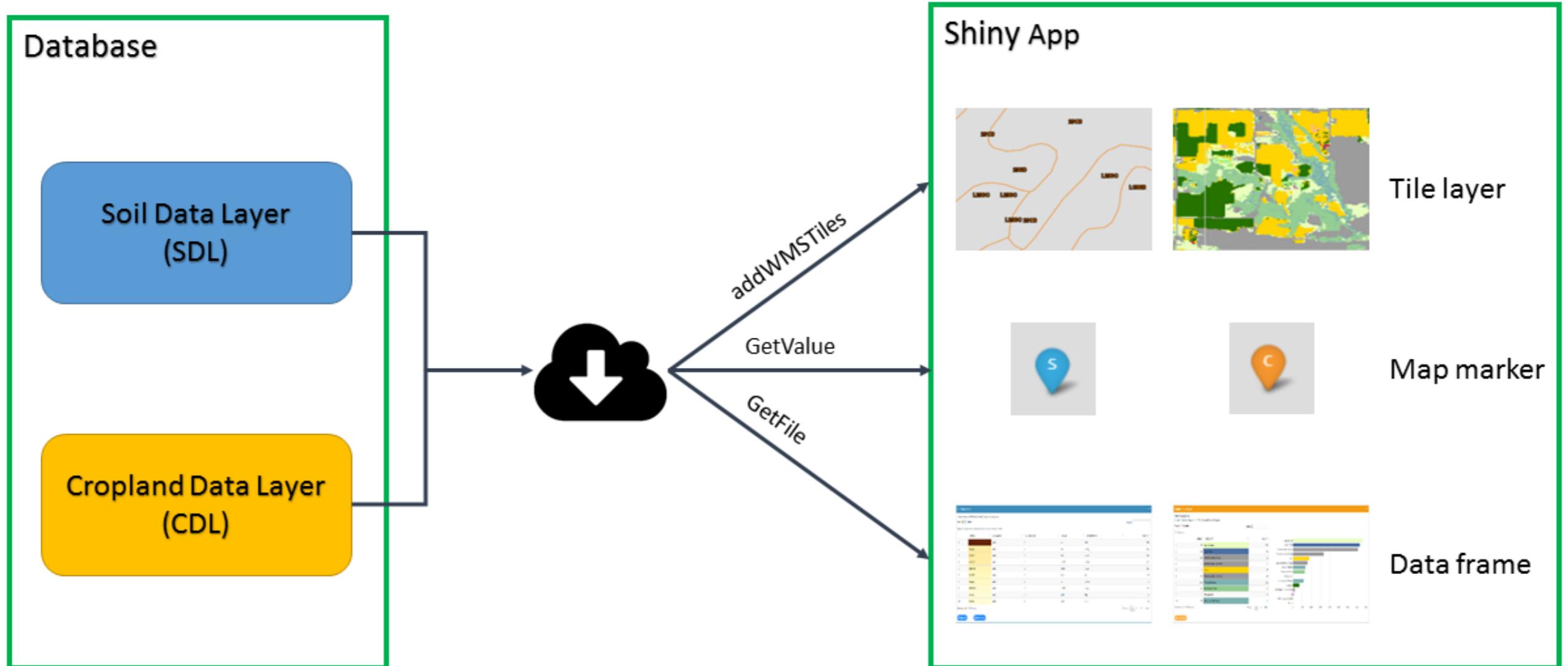
# viscover Skeleton

## Info



## Soil Data Layer

## Cropland Data Layer



Flowchart of *viscover*.

# Web Map Service

- Examples: CropScape, Soil Data Access
- returns map images in common image formats.
- input parameters: names of layers, spatial filter, projection, etc.
- in favor of commercial software, such as ESRI ArcGIS.
- `leaflet::addWMSTiles()`

# Web Geo-processing Service

- Provide access to tabular and spatial data in database.
- Request made via SQL/Extensible Markup Language (XML)
  - \* `soilDB::SDA_query(q)` - submit a query to the Soil Data Access (SDA) website in SQL.
  - \* GetCDLValue and GetCDLFile service returns XML file.

← → C ⌂  <https://nassgeodata.gmu.edu/axis2/services/CDLService/GetCDLFile?year=2009&bbox=130783,2203171,153923,2217961>

This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<ns1:GetCDLFileResponse xmlns:ns1="http://cropscape.csiss.gmu.edu/CDLService/">
  ▼<returnURL>
    https://nassgeodata.gmu.edu/webservice/nass_data_cache/CDL_2009_clip_20181017231444_53088329.tif
  </returnURL>
</ns1:GetCDLFileResponse>
```

# viscover: an R package

## ■ Installation

- \* `devtools::install_github("XiaodanLyu/viscover")`

## ■ Functions

- \* run the interactive tool: `runTool()`

- \* fetch CDL data: `GetCDLFile`, `GetCDLValue`

- \* CDL color mapping: `cdlpal`

## ■ Data

- \* CDL color scheme: `cdl.dbf`

- \* example CDL raster and soil polygon: `tile`, `poly`

# Discussion

## ■ *iNtr*

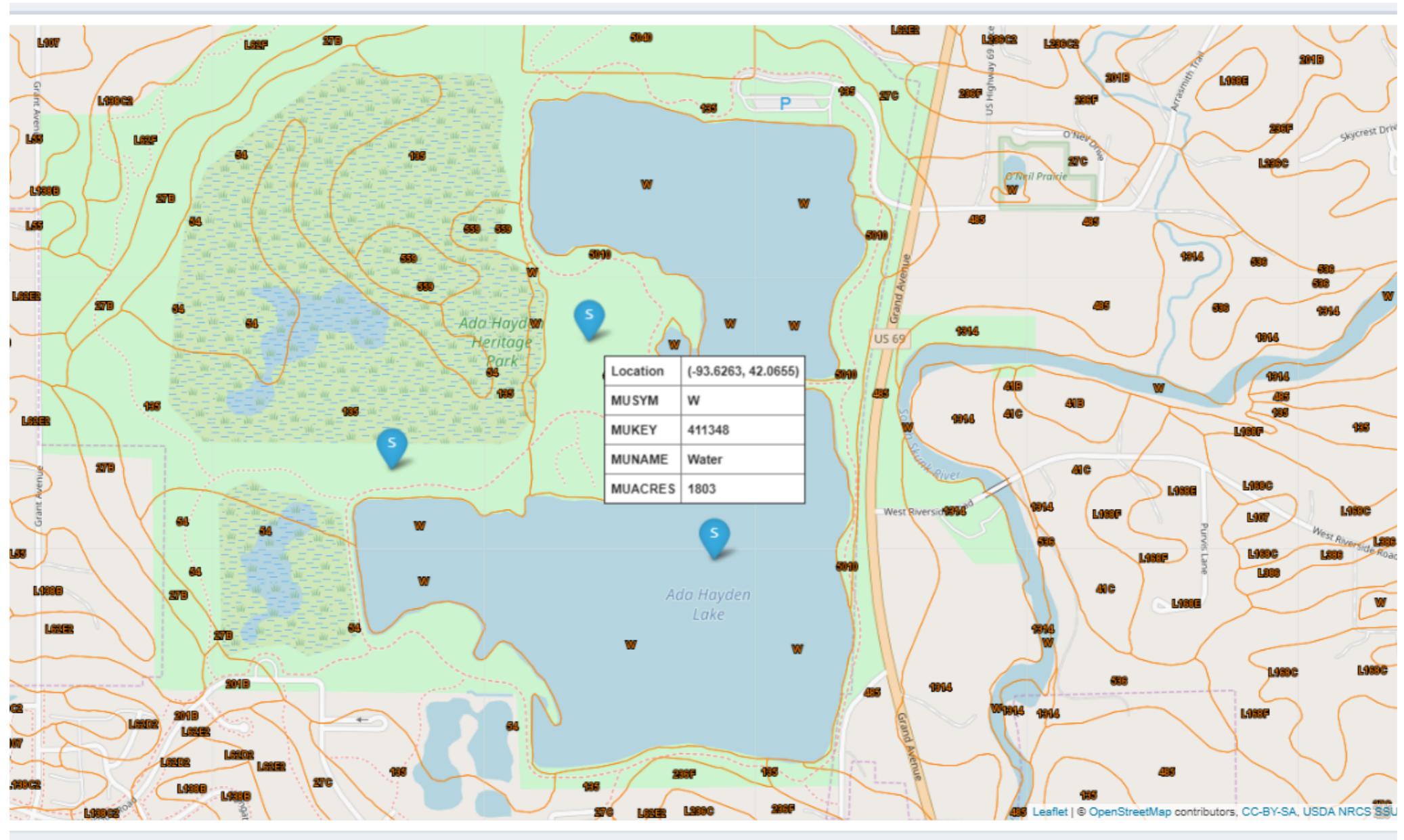
- \* used vigorously for 2015 NRI;
- \* will ultimately replace the current in-house comparison program.

## ■ *viscover*

- \* can be generalized to visualize and integrate different data sources in R.
  - Web Services seen in other official statistics: Normalized Difference Vegetation Index (NDVI), Vegetation Condition Index (VCI), National Land Cover Data (NLCD), etc.

# References

1. Groves, Robert M., and Lars Lyberg. "Total survey error: Past, present, and future." *Public opinion quarterly* 74.5 (2010): 849-879.
2. Brackstone, Gordon. "Managing data quality in a statistical agency." (2003).
3. Karr, Alan F., Ashish P. Sanil, and David L. Banks. "Data quality: A statistical perspective." *Statistical Methodology* 3.2 (2006): 137-173.
4. Williams, Matthew, and Emily Berg. "Incorporating User Input Into Optimal Constraining Procedures for Survey Estimates". *Journal of Official Statistics* 29.3 (2013): 375-396.
5. Joblin, Mitchell, and Wolfgang Mauerer. "An Interactive Survey Application for Validating Social Network Analysis Techniques." *R Journal* 8.1 (2016).
6. U.S. Department of Agriculture. 2018. Summary Report: 2015 National Resources Inventory, Natural Resources Conservation Service, Washington, DC, and Center for Survey Statistics and Methodology, Iowa State University, Ames, Iowa.
7. X. Lyu, E. J. Berg, and H. Hofmann. Empirical bayes small area prediction of sheet and rill erosion under a zero-inflated lognormal model. 2018. Manuscript submitted for publication.



**“A picture is worth a thousand words.”**