Displaying Health Data

Cases, Techniques, Solutions

Colloquium + Live-Webcast + Recording
Medical Sciences Building (MBS) 160
University of Victoria
November 28 – 30 , 1 – 3 pm PST

University of Victoria
INSTITUTE ON AGING & LIFELONG HEALTH
island health
British Columbia BSERVATORY Population & Public Health
IALSA
Health Sciences Center
UCF
UNIVERSITY OF CENTRAL FLORIDA

Displaying Health Data

Cases, Techniques, Solutions

Health Data

DAY 1
2018-11-28
Wednesday

13:00 Transactional data of Island Health: How patients vote with their feet
Dr. Ken Moselle (Island Health) and Dr. Andriy Koval (BC Observatory, UCF)

14:00 Visualizing logistic regression with the "coloring book" technique: A study in ggplot2
Dr. Andriy Koval (BC Observatory for Population and Public Health, UCF)

Substance Use

DAY 2
2018-11-29
Thursday

13:00 Nuances of information sharing and data display in a mobile application for students with substance use disorder
Dr. Barbara (Basia) Andraka-Christou (University of Central Florida)

14:00 Optimizing public health surveillance through reproducible reporting: Response to opioid crisis on Vancouver Island
Shannon Tracey (University of Victoria) and Maritia Gully (Island Health)

Pipelines & Dashboards

DAY 3
2018-11-30
Friday

13:00 Building pipelines and dashboards for practitioners: Mobilizing knowledge with reproducible reporting
Dr. Will Beasley (University of Oklahoma Health Sciences Center)

14:00 Constructing workflows for reproducible analytics: Suppressing small counts for provincial chronic disease dashboard
Dr. Andriy Koval (BC Observatory, UCF) and Anthony Leamon (Island Health)

INSTITUTE ON AGING & LIFELONG HEALTH

University of Victoria

island health

British Columbia BSERVATORY Population & Public Health

IALSA

Health Sciences Center

UCF UNIVERSITY OF CENTRAL FLORIDA

Building pipelines and dashboards for practitioners

Knowledge mobilization with reproducible reporting

Will Beasley

The University of Oklahoma Health Sciences Center

Please email questions to
aging@uvic.ca

Access lecture slides from

github.com/dss-ialh/displaying-health-data

Clone the repo to reproduce examples

Building Pipelines and Dashboards for Practitioners:

Mobilizing knowledge with reproducible reporting

Will Beasley, Geneva Marshall, & David Bard
University of Oklahoma Health Science Center
Biomedical and Behavioral Health Core

Common Roles for the BBMC

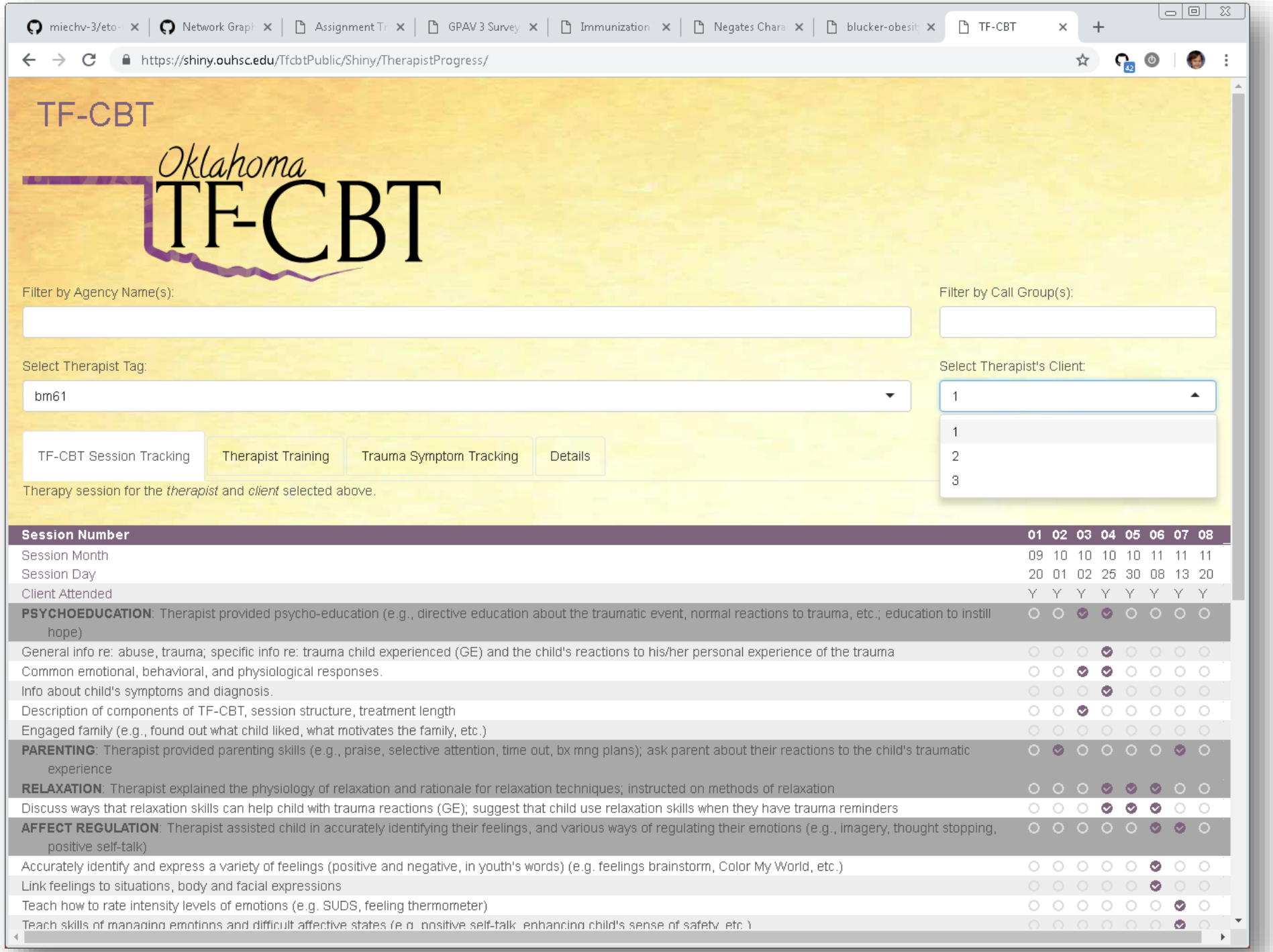
domain expertise in a few areas

incorporation of data sources from multiple agencies

statistical analysis for academic publications

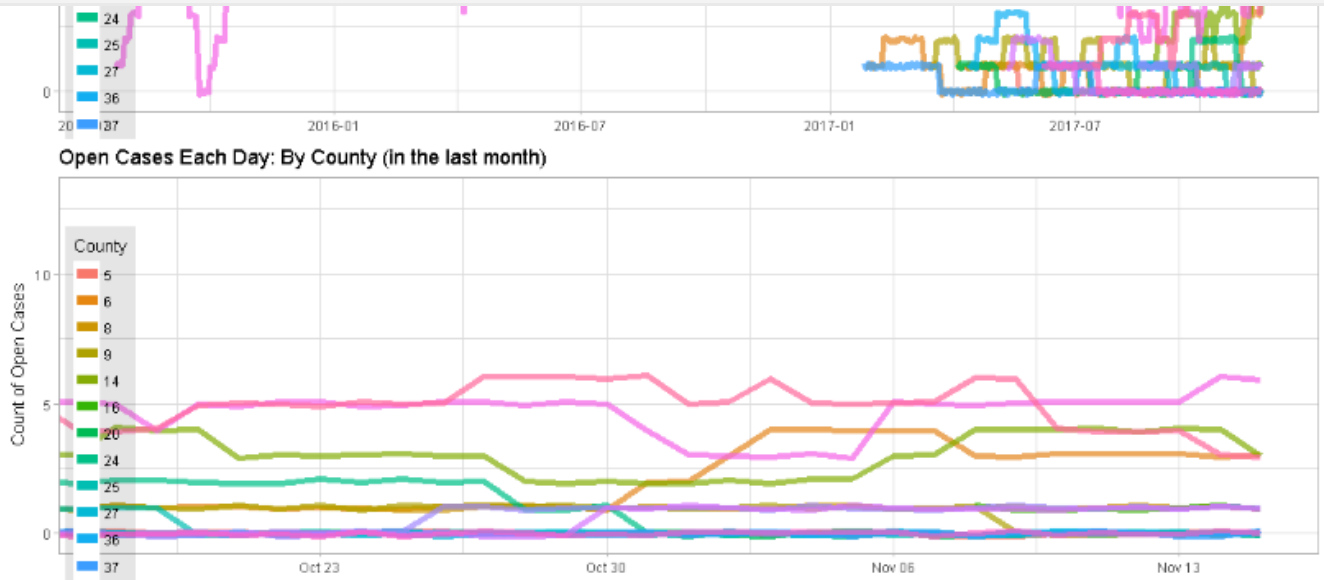
construct information displays for nontechnical audiences

Example 1: history of therapist training



Example 2: availability of specialized CPS workers

1 Summary
2 State-level View
3 Supervisor-level View
4 County-level View
4.0.1 Current
4.0.2 Previous
5 Worker-level View
6 Referral-level View
7 Session Information



5 Worker-level View

5.0.1 Current

Each row represents one ISS worker. The 'capacity by region' field squishes together the capacities for the five regions, separated by commas.

iss worker	region(s)	county(ies)	stop date	active worker	capacity by region	capacity max	assignments open [total]	open records	has capacity
95 [redacted] {Blaine/Canadian/Custer}	1;	6;9;20	-	Yes	1,-,-,-,-	1	1 [43]	1658	no
488 [redacted] {Canadian}	1;	9	-	Yes	0,-,-,-,-	0	0 [2]	-	no
749 [redacted] {Blaine/Canadian/Garfield/Grant/Kay/Kingfisher/Logan/Noble/Payne}	1;	6;9;24;27;36;37;42;52;60	-	Yes	3,-,-,-,-	3	3 [15]	1622; 1625; 1708	no

Example 3:
data collection quality

3 Smells

A *smell* test won't validate a given record (like the rules above), but it will make sure the dataset overall smells correct.

3.1 Smell Summary

Smell Name	Priority	Proportion of...	Legal Boundaries	Value	Pass
female majority	**1**	**female respondents.**	**[0.80, 1.00]**	**0.0000**	**FALSE**
age proportion infant	2	index children under 2 months old.	[0.05, 1.00]	0.1660	TRUE
age proportion toddler	2	index children between 2 & 12 months old.	[0.20, 1.00]	0.2510	TRUE
proportion black	**2**	**Around 4th of the survey population is black**	**[0.00, 0.30]**	**0.3166**	**FALSE**
proportion children at home	2	Most participants have children living at home most of the time.	[0.50, 1.00]	0.8803	TRUE
proportion currently pregnant	2	More than 2/3rd of the participants are not currently pregnant.	[0.80, 1.00]	0.8672	TRUE
proportion developmental disabilities	2	Less than 10% of the children have developmental disabilities.	[0.00, 0.10]	0.0714	TRUE
proportion food stamps services	2	A little less than half the population is on food stamps.	[0.35, 0.50]	0.4981	TRUE
proportion hispanic	**2**	**A little less than half the survey population is Hispanic.**	**[0.30, 0.40]**	**0.2857**	**FALSE**
proportion household children	2	Around half of our survey population has households with one child.	[0.40, 0.50]	0.4942	TRUE

GPAV 3 Survey Validation

Date: 2018-11-06 10:46:50

This report identifies violations in the GPAV 3 Community Survey dataset, aka (Giving Parents A Voice -MIECHV 3).

1 Set Up

1.1 Instructions | 1.2 Execute Rules | 1.3 Execute Smells | 1.4 Inactive Rules | 1.5 Inactive Smells

1. Work through each line in the table to discover & correct entry mistakes in the database. Be aware that a single mistake may manifest in multiple rows; fixing one value may clear several rows.
2. Click the `record id` value in a row to be taken to the participant's Event Grid in REDCap (*i.e.*, the 'stop light page').
3. Double-check that your manual corrections in REDCap are reflected in data-public/derived/gpav/gpav-3-violation-rule.csv. Be careful not to move this file to somewhere unsafe.
4. To help discover the dataflow and problematic location, use the
   - project's codebook in REDCap,
   - gpav-3-arch file, and
   - validation-gpav-3 report.
5. If you create a new rule or smell check, make sure the error message won't reveal any PHI.

2 Rules

A *rule* is very exact. Each record is examined, and determined if it passes each specific rule.

2.1 Rule Detail | 2.2 Rule Summary

Violations at 2018-11-06 10:47:04

	check name	record id	data collector	error message	priority	instrument	interview date
				All		All	
1	parent pro services	255-24	8	ParentPRO home-based services status is missing. (The question is 'Are you currently participating in parentPRO home-based services?')	1	participant demographics	2018-09-05
2	parent pro services	255-63	8	ParentPRO home-based services status is missing. (The question is 'Are you currently participating in parentPRO home-based services?')	1	participant demographics	2018-10-09
3	Parentpro service	255-24	8	ParentPRO service status is missing when recruit source is present. If recruit source	1	participant	2018-09-05

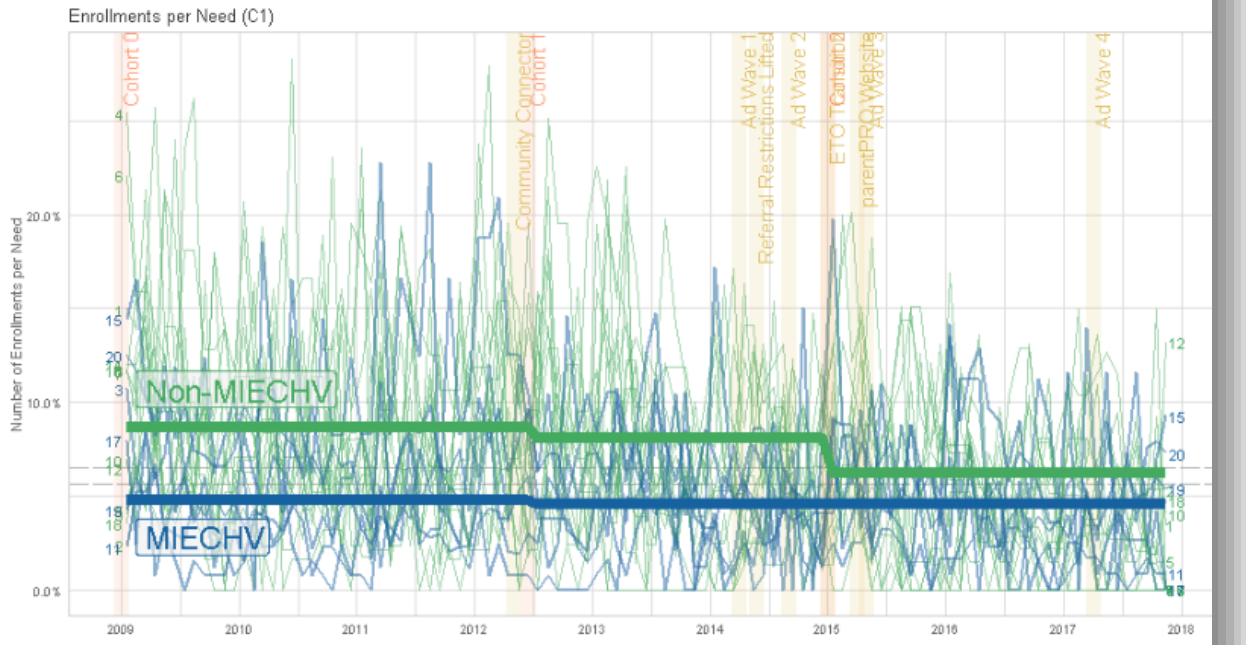
Example 4: MCMC HPDs are updated as collection progresses

Example 5: semi-annual reports with multilevel longitudinal models

completed enrollments (not significant for the drop between 0 and 1, but significant for the drop between 0 and 2. Coefficients *1d* and *1e* are greater than 1.00, but not significant; a significant coefficient *would have* indicated that the six MIECHV regions are comparatively outgaining the fourteen comparison regions between cohorts 0 and 2. However, *1b* × *1d* produces a value under 1.00 (and *1c* × *1e* do too), indicating MIECHV regions observed a decline in overall enrollments (but a much less steep decline than comparison regions). This pattern is evident in Figure zzz; notice how the green line descends sharply during the nine year period, while the blue line is essentially flat. The groups' enrollment rates are almost equal by 2015.

The esimates from Models 2 and 3 corresponded closely to Model 1, suggesting that the three operationalizations of the MIECHV treatment are comparable for completed enrollments. The only notable difference involves coefficient *g*. The Model 2 and 3 estimates are about 5%, at which point coefficients become significant; this suggests MIECHV regions did relatively better than non-MIECHV regions between cohorts 1 and 2.

		MIECHV Contracted			MIECHV Staffed			Has Connector	
			Change	p		Change	p	Change	p
0 - NvM	:	1a	0.57	<.001	2a	0.62	<.001	3a 0.68	<.001
0v1 - N	:	1b	0.93	.235	2b	0.94	.181	3b 0.97	.533
0v2 - N	:	1c	0.71	<.001	2c	0.76	<.001	3c 0.76	<.001
0v1 - NvM	:	1d	1.01	.925	2d	1.00	.978	3d 0.89	.088
0v2 - NvM	:	1e	1.29	.103	2e	1.17	.003	3e 1.18	.002
1v2 - N	:	1f	1.05	.287	2f	0.99	.744	3f 0.96	.276
1v2 - NvM	:	1g	1.23	<.001	2g	1.26	<.001	3g 1.31	<.001



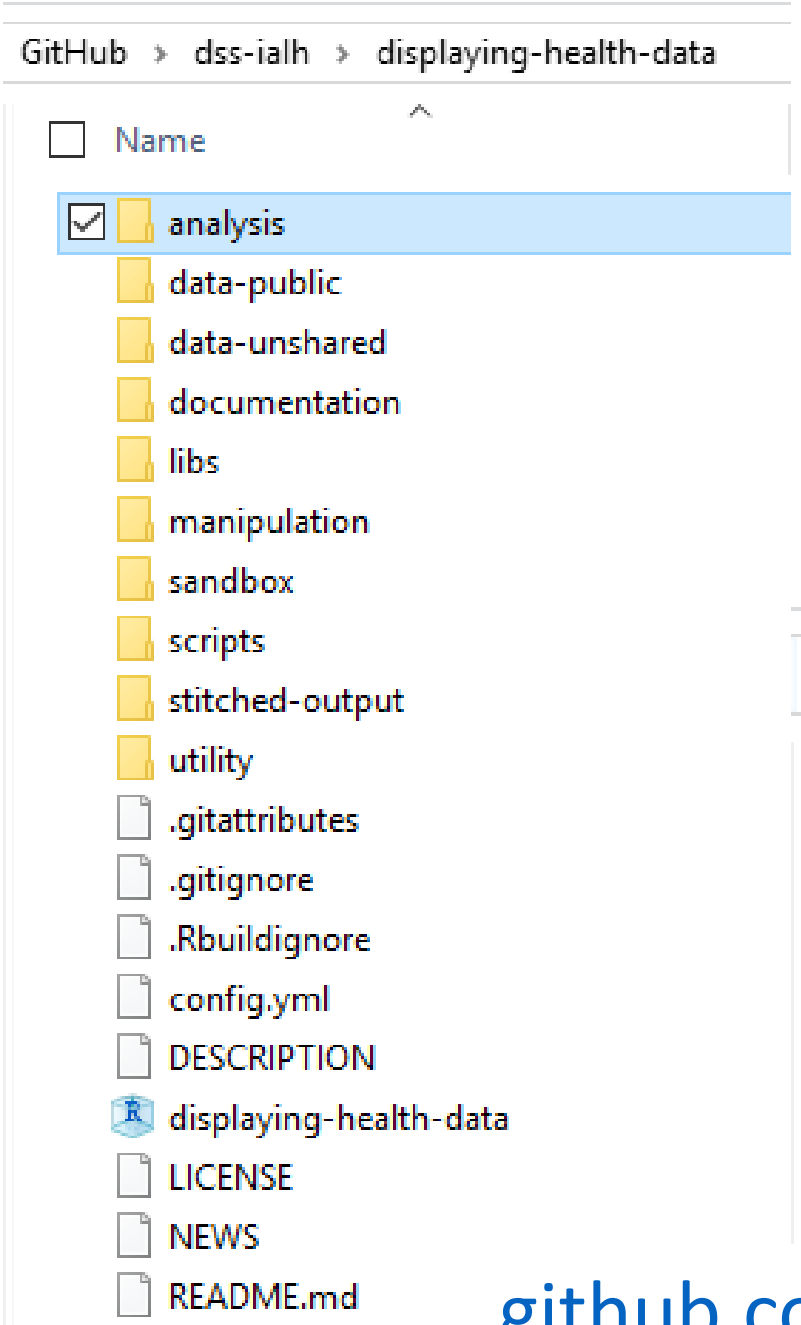
Response variable: `EnrollCompletedCount` ; *Offset variable*: `InfantNeedCountForTime`

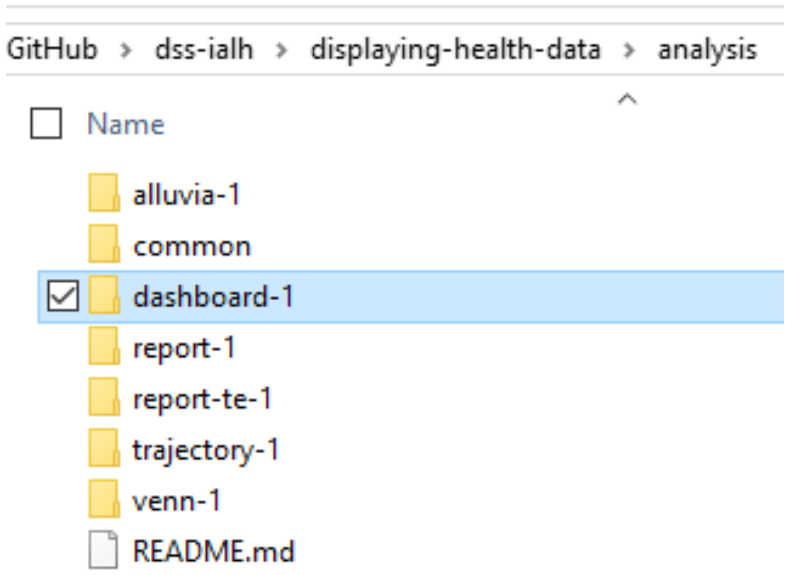
dashboards vs reports

- Blur the distinction

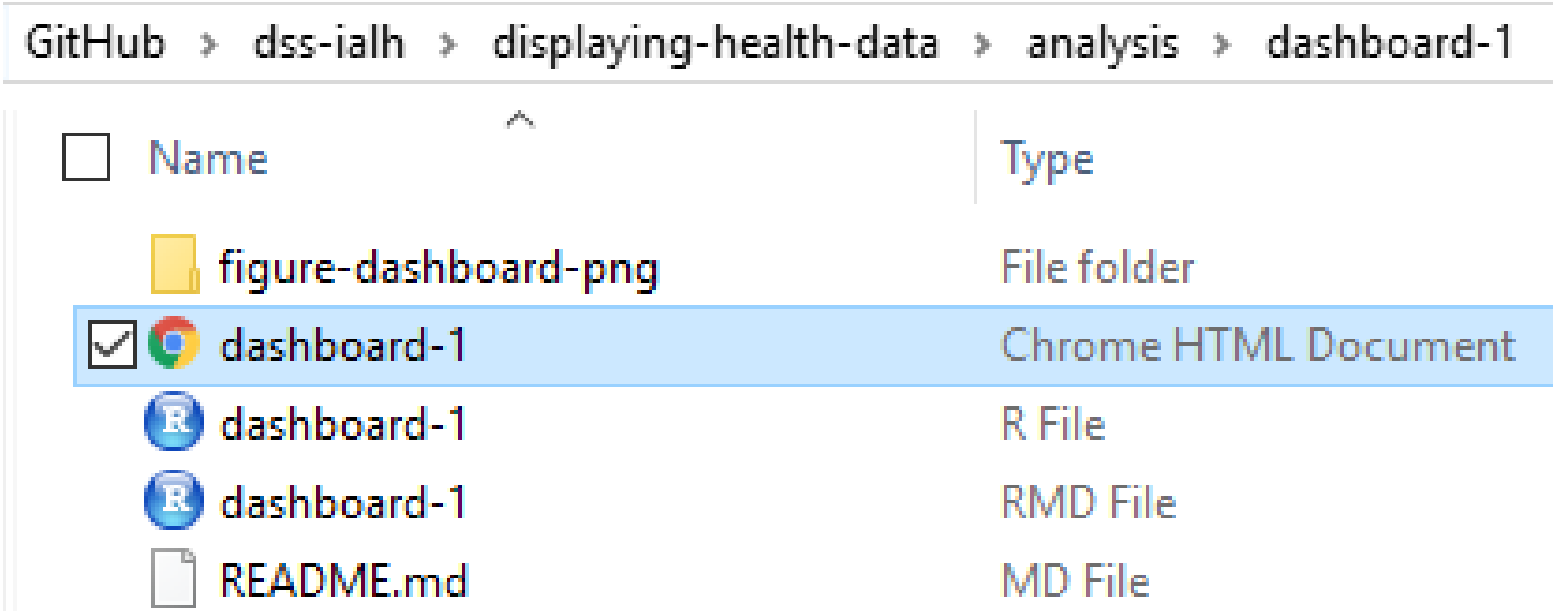
- Try to make your reports more like dashboards, so they
  1. Are automated,
  2. Are frequently updated & consumed (especially early in the project),
  3. Actively flag errors, and
  4. Are easy and fun to jump into

- Try to make your dashboards more like reports, so they
  1. Provide context and are self-explanatory,
  2. Are portable and archivable, and
  3. Contain statistical analyses that encourage sophisticated reasoning.





Fictional, Reproducible example



github.com/dss-ialh/displaying-health-data

Pipeline (available at github.com/dss-ialh/displaying-health-data)

Raw CSVs

The external data's first appearance in our system.

`data-public/raw/`

Central Database

The inspected & loaded data into tables.

`data-public/derived/`

Analysis-Ready Rectangles

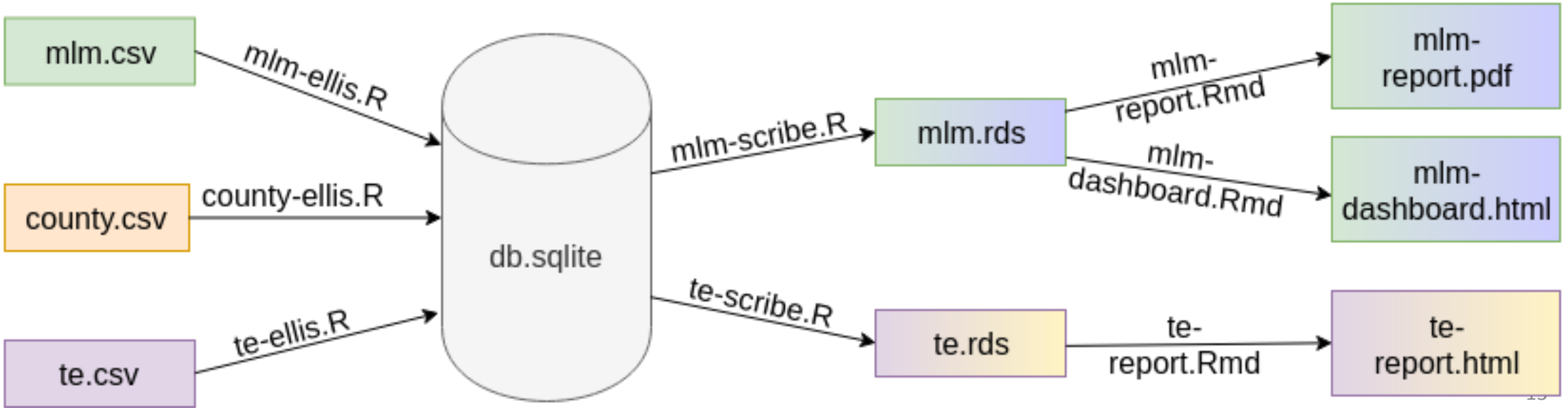
The joined & groomed datasets easily digestible.

`data-public/derived/`

Analysis

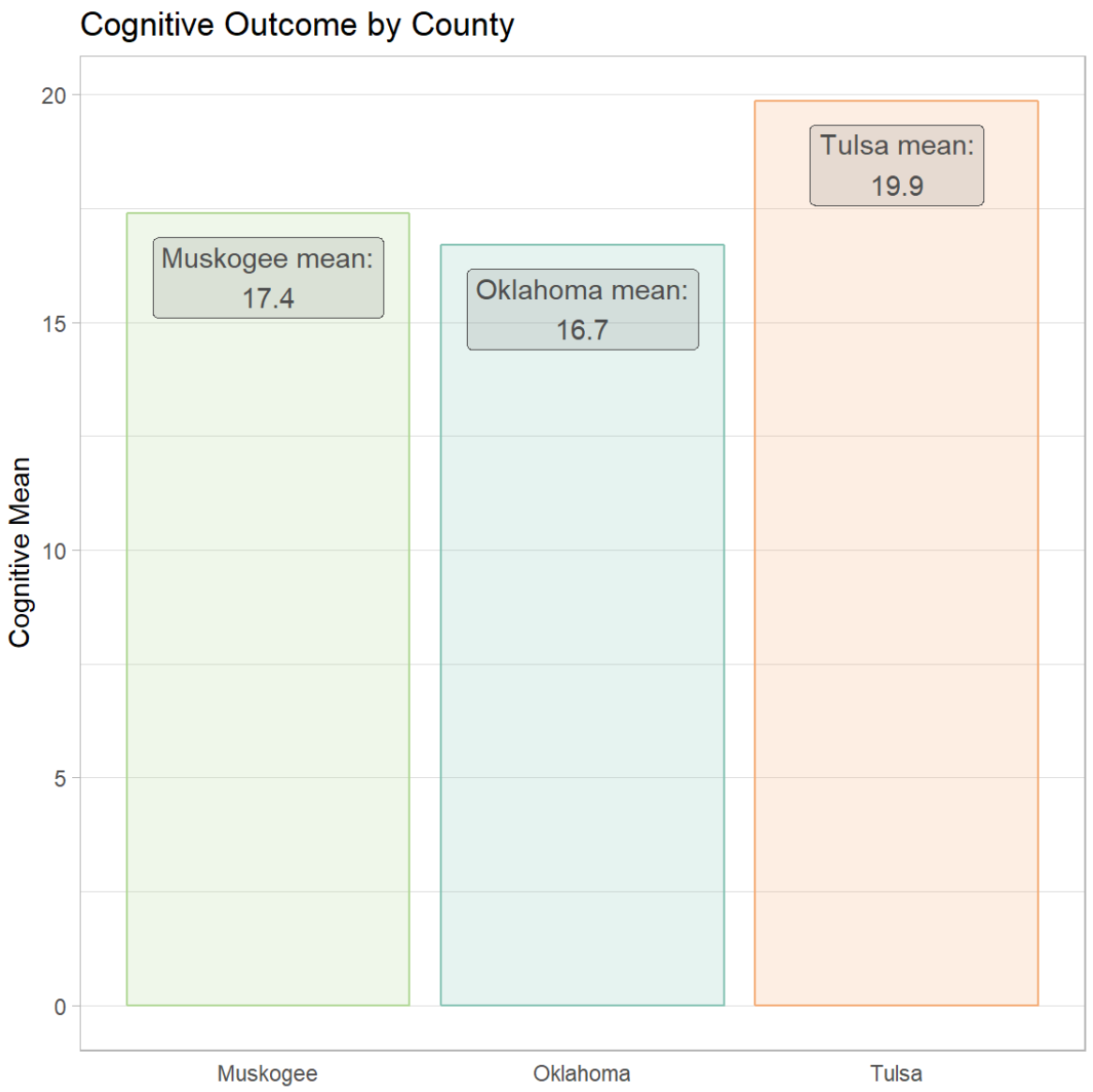
Products for internal & external audiences,

`analysis/`



Fictional scenario

- We partner with the government to monitor the cognitive trajectories of patients over time.
- Recruited 200 seniors from three county programs.
- Dashboard is tailored to providers in Tulsa



In the real study…

The nurses create three 6-month mini-research studies.

They develop a manageable change to the services the provide, and see if their mini-intervention improves an outcome they've chosen.

One real intervention developed a thorough script of welcoming new clients. A second intervention was sending text messages to clients to reduce no-shows. A third intervention developed guidelines for discussing traumatic ACEs in their past.

Starts with 2-day workshop to develop the ideas.

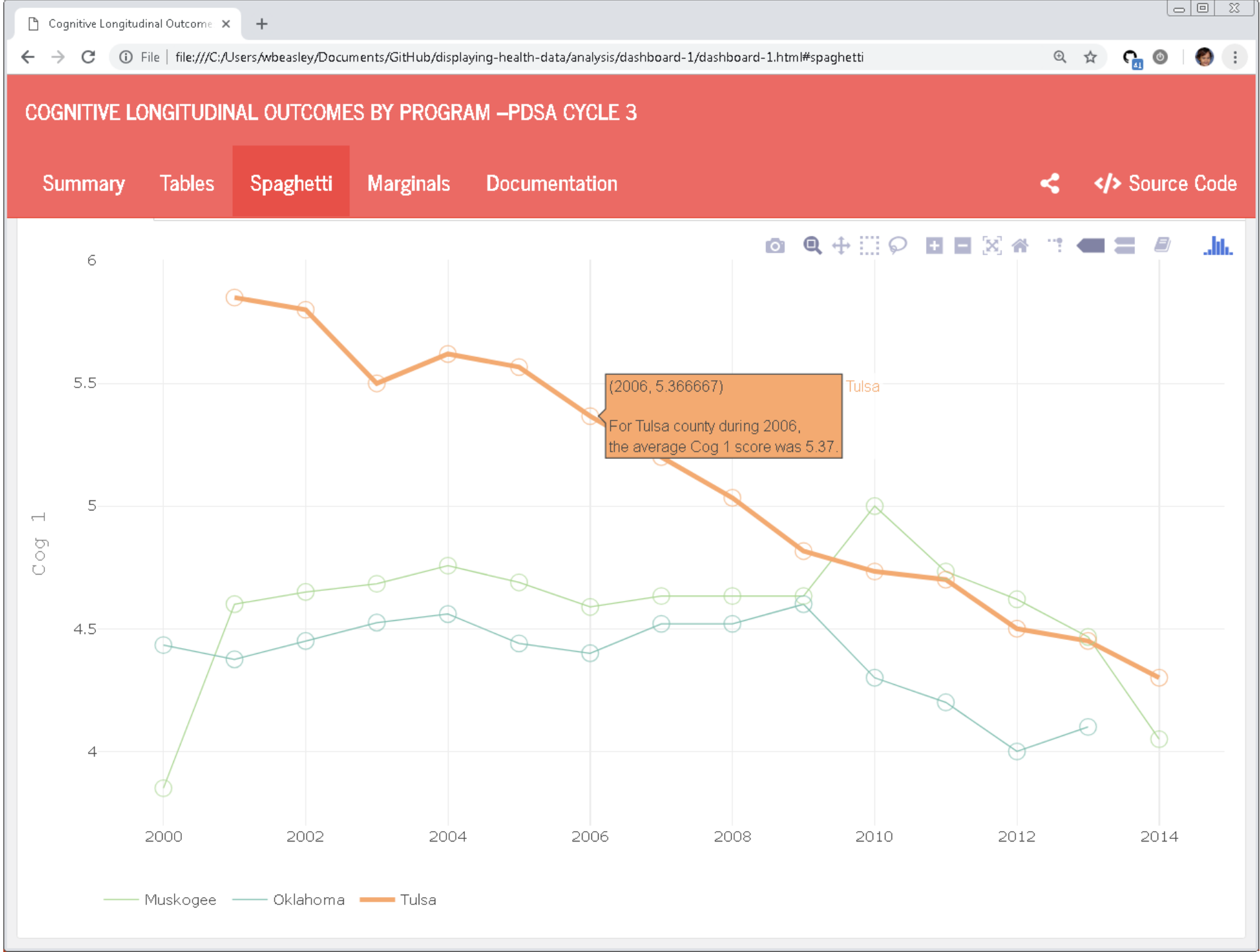
Data Glimpse

(start demo)

```
> ds %>% dplyr::glimpse(50)
Observations: 200
Variables: 21
$ subject_wave_id <int> 1, 2, 3, 4, 5, 6, 7, ...
$ subject_id      <int> 1001, 1001, 1001, 100...
$ county_id       <int> 51, 51, 51, 51, 51, 5...
$ gender_id       <dbl> 2, 2, 2, 2, 2, 2, 2, ...
$ race            <chr> "Native Hawaiian or O...
$ ethnicity       <chr> "Hispanic or Latino",...
$ county          <chr> "Muskogee", "Muskogee...
$ wave_id         <int> 1, 2, 3, 4, 5, 6, 7, ...
$ year            <int> 2000, 2001, 2002, 200...
$ date_at_visit   <date> 2000-04-25, 2001-02-...
$ age             <int> 67, 68, 69, 70, 71, 7...
$ age_cut_4       <chr> "60s", "60s", "60s", ...
$ age_80_plus     <int> 0, 0, 0, 0, 0, 0, 0, ...
$ int_factor_1    <dbl> 8.895, 8.895, 8.895, ...
$ slope_factor_1  <dbl> -0.029, -0.029, -0.02...
$ cog_1           <dbl> 3.3, 3.5, 3.5, 3.4, 3...
$ cog_2           <dbl> 4.6, 4.6, 4.2, 4.7, 4...
$ cog_3           <dbl> 5.3, 5.3, 5.2, 5.2, 5...
$ phys_1          <dbl> 2.8, 2.9, 2.8, 3.1, 3...
$ phys_2          <dbl> 3.9, 3.5, 3.9, 3.9, 3...
$ phys_3          <dbl> 0.5, 0.0, 0.7, 1.3, 2...
```

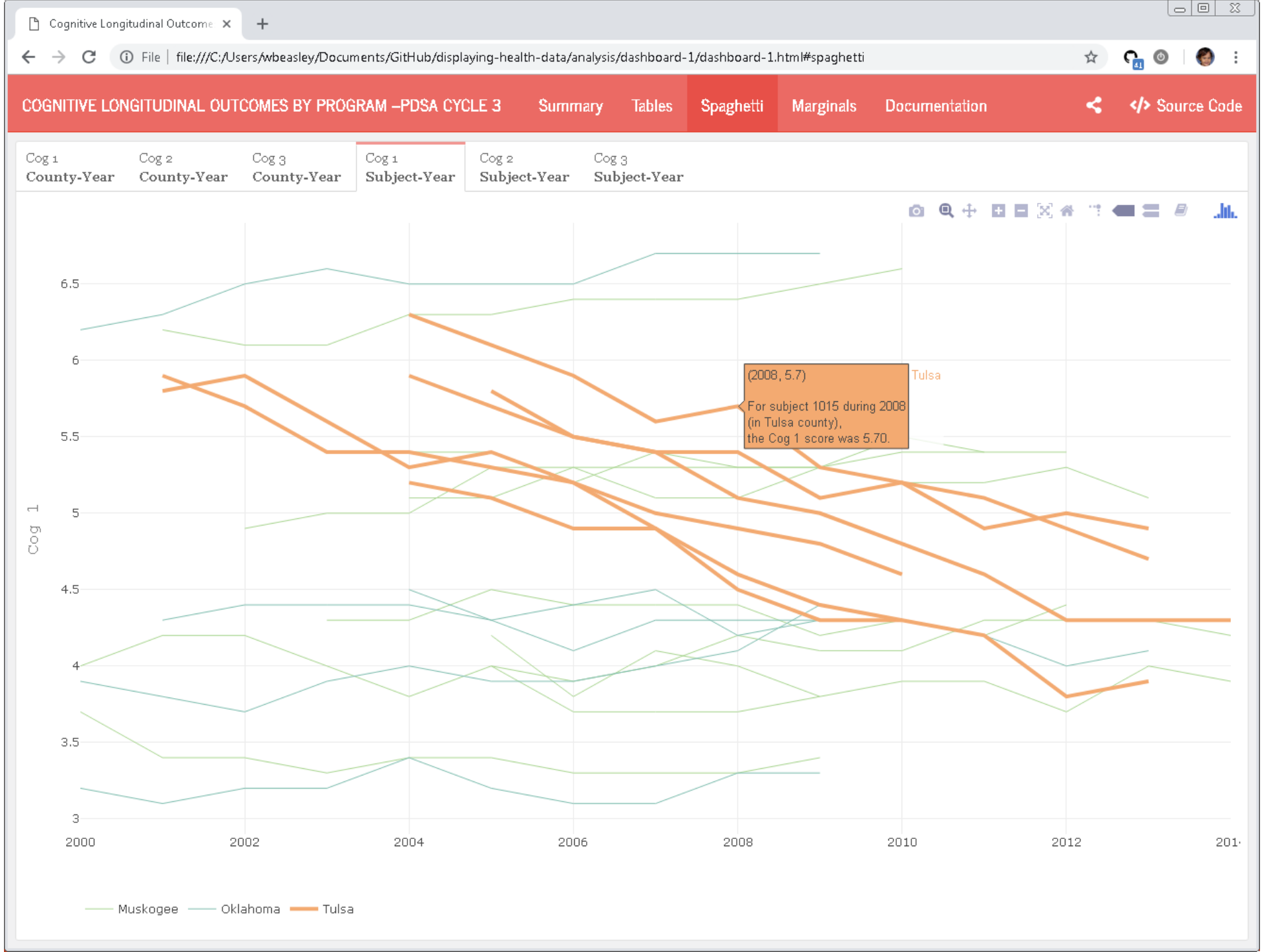
Program directors need to monitor the focal group over time.

De-emphasized programs provide meaningful reference points.

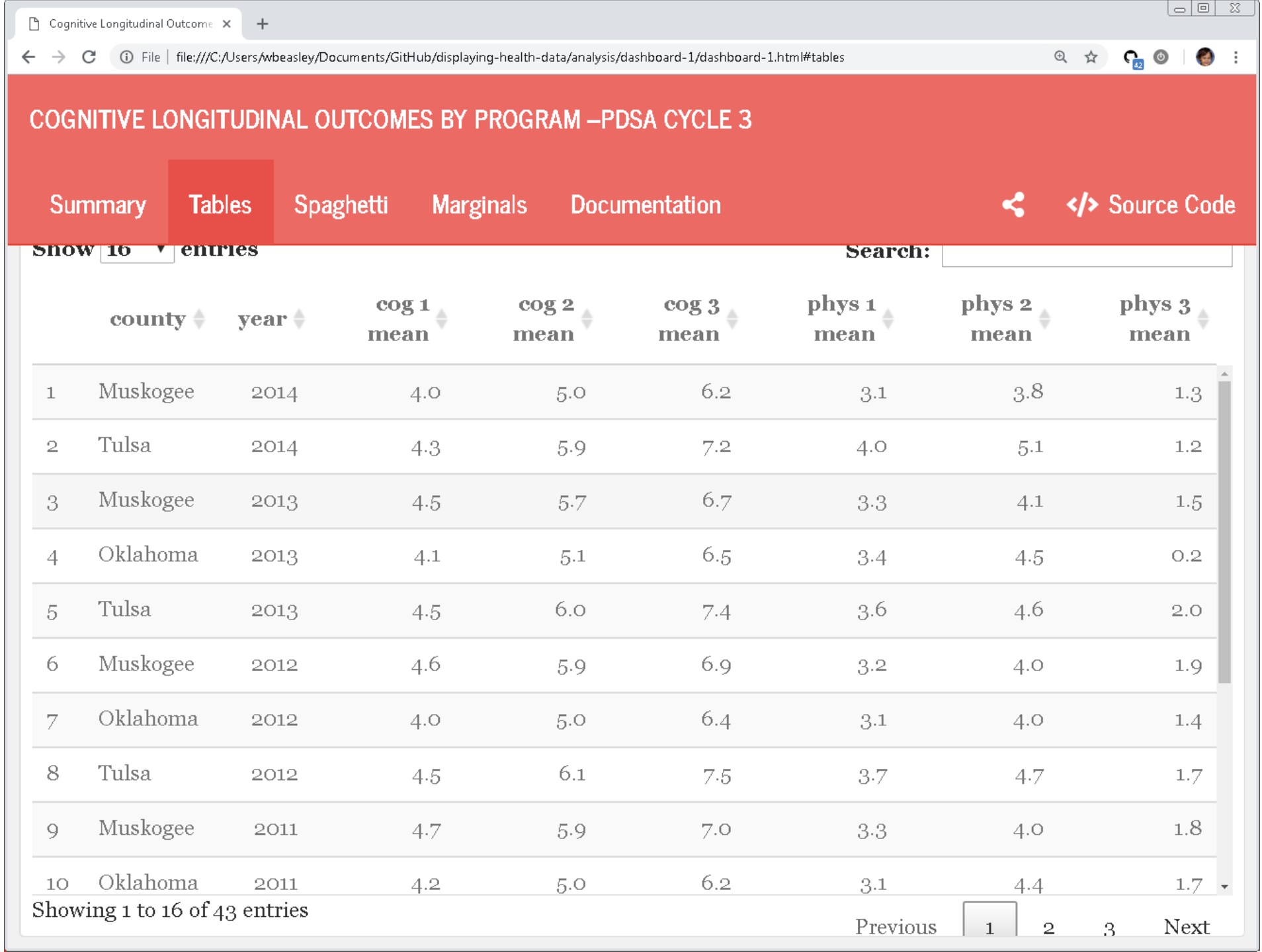


We also need to see the progress of individuals

Tulsa county trajectories are in orange. The others are primarily for context.

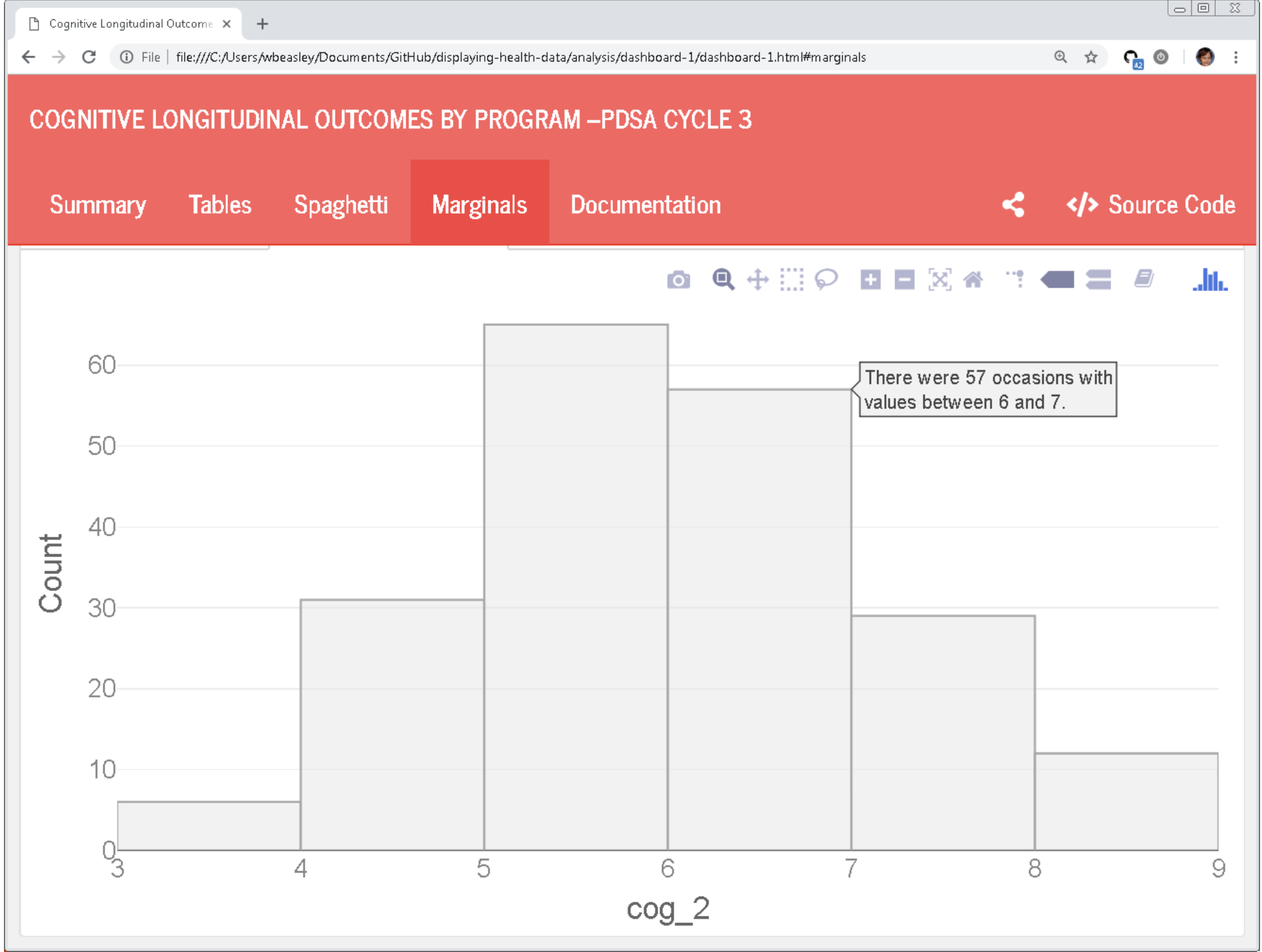


Some nurses were more comfortable starting with tables than graphs.



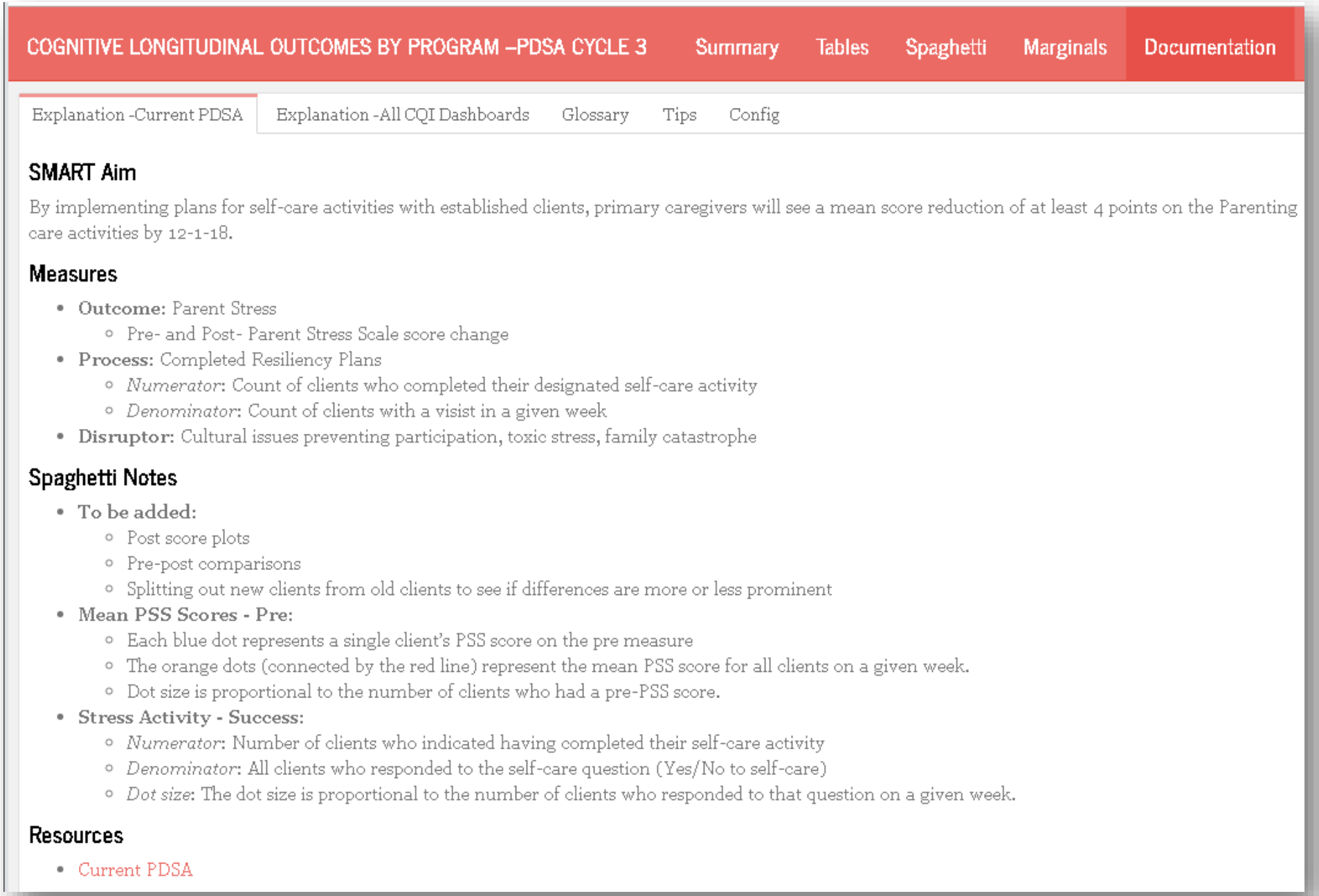
	county	year	cog 1 mean	cog 2 mean	cog 3 mean	phys 1 mean	phys 2 mean	phys 3 mean
1	Muskogee	2014	4.0	5.0	6.2	3.1	3.8	1.3
2	Tulsa	2014	4.3	5.9	7.2	4.0	5.1	1.2
3	Muskogee	2013	4.5	5.7	6.7	3.3	4.1	1.5
4	Oklahoma	2013	4.1	5.1	6.5	3.4	4.5	0.2
5	Tulsa	2013	4.5	6.0	7.4	3.6	4.6	2.0
6	Muskogee	2012	4.6	5.9	6.9	3.2	4.0	1.9
7	Oklahoma	2012	4.0	5.0	6.4	3.1	4.0	1.4
8	Tulsa	2012	4.5	6.1	7.5	3.7	4.7	1.7
9	Muskogee	2011	4.7	5.9	7.0	3.3	4.0	1.8
10	Oklahoma	2011	4.2	5.0	6.2	3.1	4.4	1.7

Histograms are particularly benefited from full-sentence hover text.



Each dashboard had several tabs of documentation.

Our quality varied across dashboards, and this is a representative example.



Lessons Reinforced

1. The providers have incredible ideas for improving services delivered. It's worth our time to facilitate and translate these ideas.

2. Implementing their tx & measurements ideas increases their engagement.

3. Incorporate the providers in the dashboard design

4. Math phobia is real, and it might be hard for your developers to related to.

5. Writing the scribes is difficult (ie, grooms and reshapes multiple analysis-agnostic data tables into a single analysis-specific rectangle).

Lessons Learned

1. This particular group of nurses preferred a top-down entry into their data. Most scientists prefer bottom-up. Here are some examples of how we changed our presentation.

2. Hovering text provides an alternative entry into understanding the graph.

3. Incorporate the providers in the dashboard design

4. Monthly meetings are not frequent enough at the beginning

Suggestions for dashboard improvements?

- Specific graphs
- Presentation order
- Verbal documentation
- My approach/attitude towards the information
- My approach/attitude towards the consumers
- Feature to emphasis or de-emphasize

Pipeline (available at github.com/dss-ialh/displaying-health-data)

Raw CSVs

The external data's first appearance in our system.

`data-public/raw/`

Central Database

The inspected & loaded data into tables.

`data-public/derived/`

Analysis-Ready Rectangles

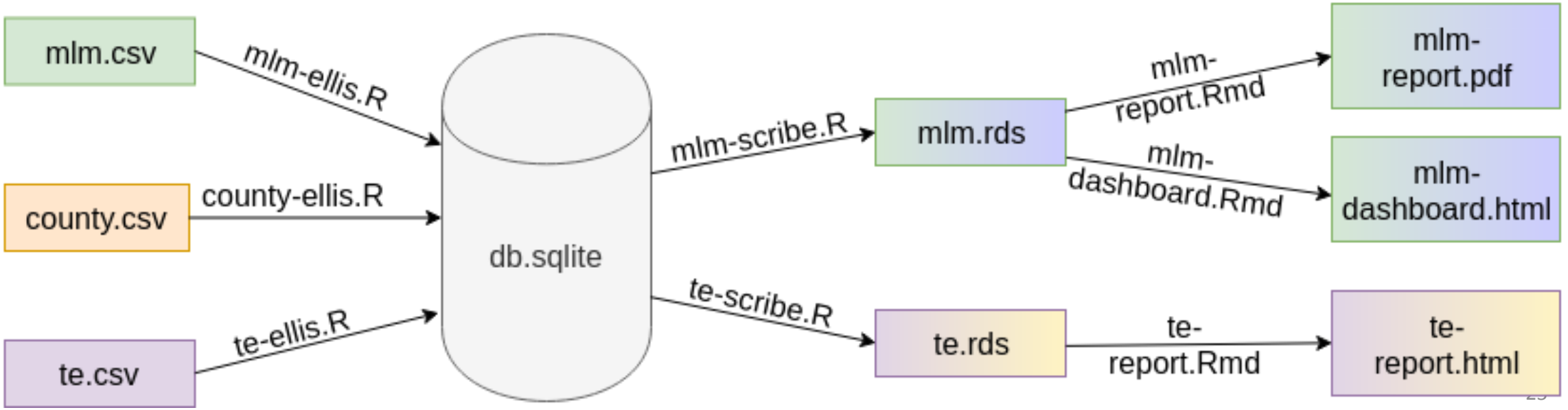
The joined & groomed datasets easily digestible.

`data-public/derived/`

Analysis

Products for internal & external audiences,

`analysis/`



Pipeline

Raw CSVs

The external data's first appearance in our system.

`data-public/raw/`

Central Database

The inspected & loaded data into tables.

`data-public/derived/`

Analysis-Ready Rectangles

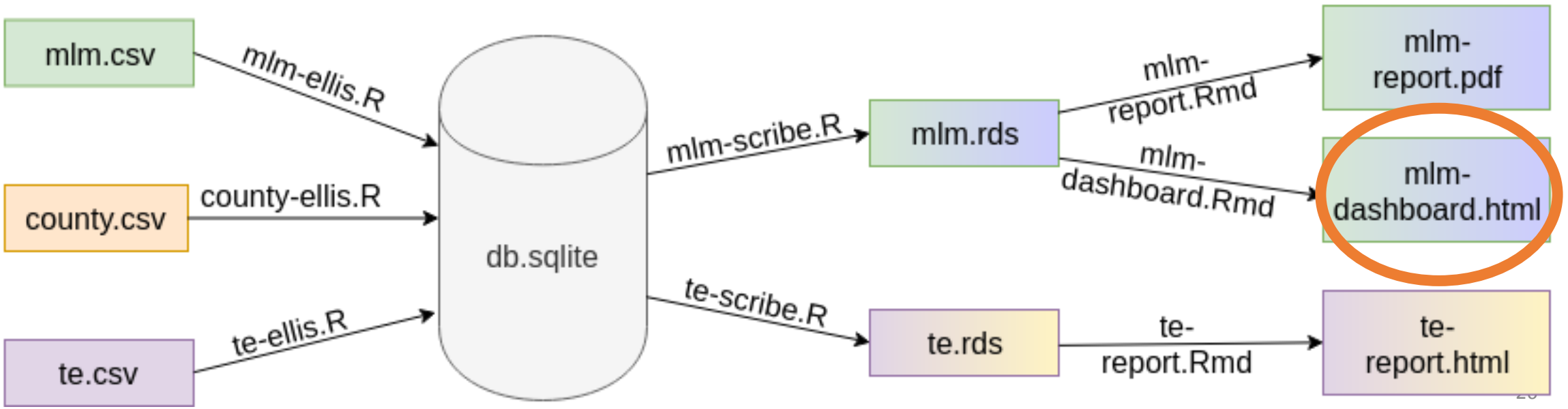
The joined & groomed datasets easily digestible.

`data-public/derived/`

Analysis

Products for internal & external audiences,

`analysis/`

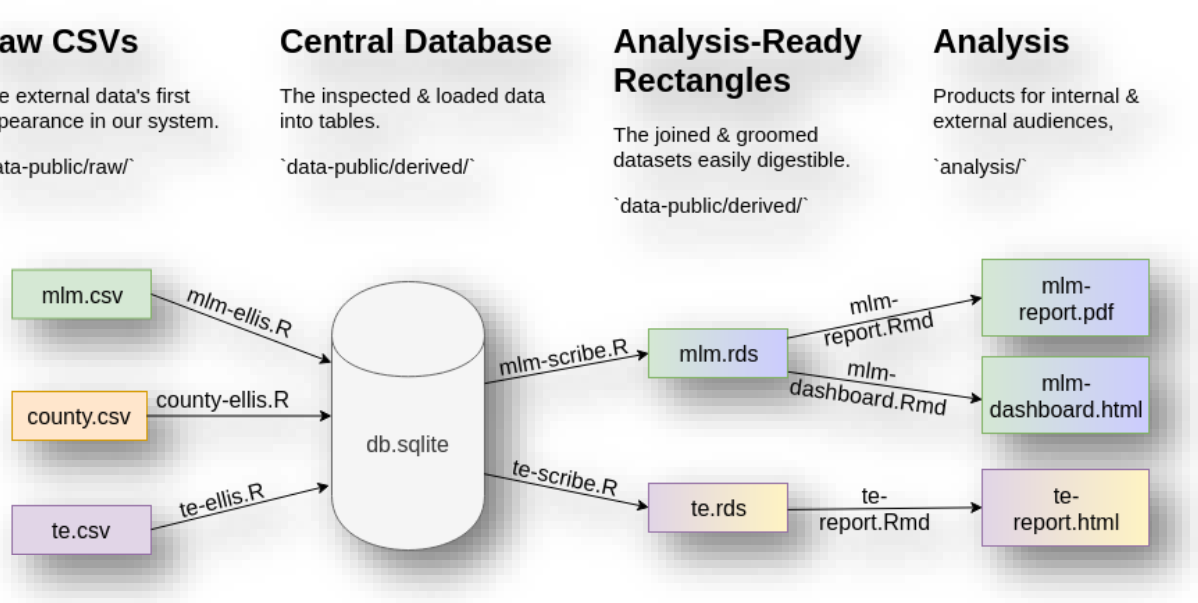


Benefits

1. Scaling
    1. more data sources
    2. multiple analysts
    3. multiple reports

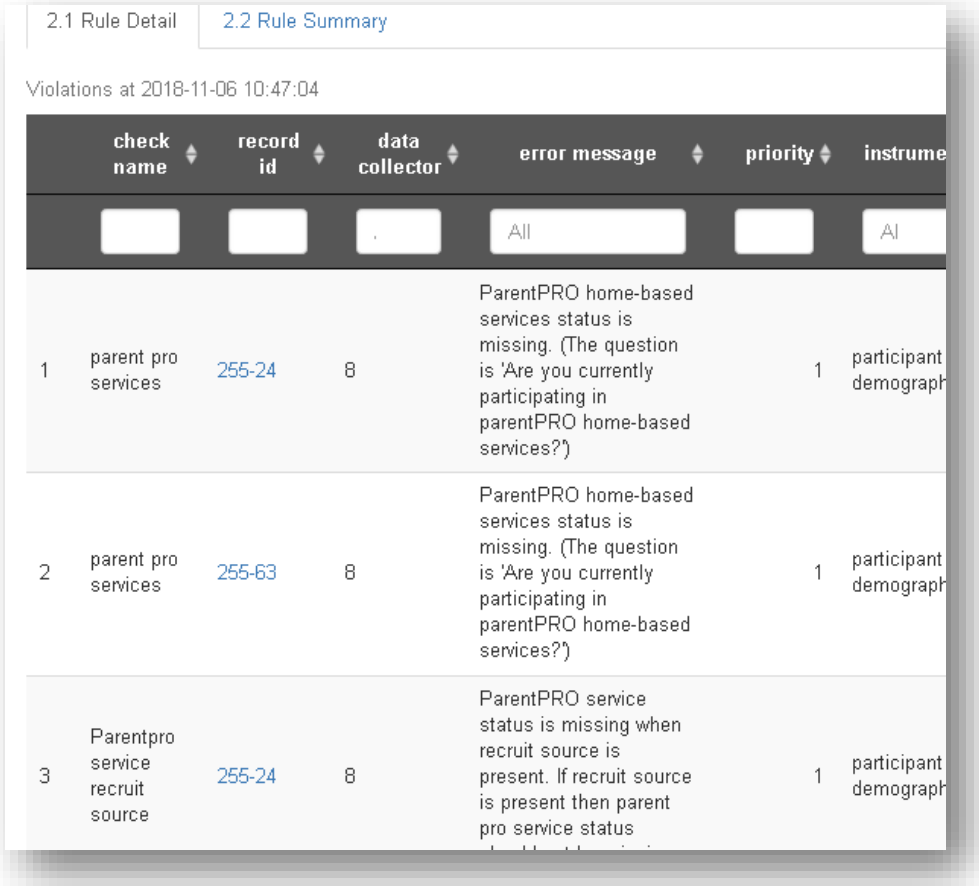
2. Hide PHI for as long as possible

3. Leverage database performance



Deployment Options

1. Server-rendered vs Client-rendered
2. Interactive vs static
3. Stage to redact PHI
4. Reconnect to PHI record using authentication
5. human vs cronjob



2.1 Rule Detail | 2.2 Rule Summary

Violations at 2018-11-06 10:47:04

	check name	record id	data collector	error message	priority	instrume
				All		Al
1	parent pro services	255-24	8	ParentPRO home-based services status is missing. (The question is 'Are you currently participating in parentPRO home-based services?')	1	participant demograph
2	parent pro services	255-63	8	ParentPRO home-based services status is missing. (The question is 'Are you currently participating in parentPRO home-based services?')	1	participant demograph
3	Parentpro service recruit source	255-24	8	ParentPRO service status is missing when recruit source is present. If recruit source is present then parent pro service status	1	participant demograph

Providers in Oklahoma

1. **Oklahoma State Department of Health**– Persephone Starks, John Delara, Beth Martin
2. **Children First Tulsa** – LouAnn Beuke
3. **LCDA** – Patty Demoraes-Huffhine
4. **Oklahoma City County Health Department** – Denise Howard, Diane Sammons, Sally Dixon
5. **OKC Public Schools** – Yolanda Lucero, Kethzia Njikam
6. **PAT Bethany Public Schools** – Mindy Turner, Jem Balderas
7. **PAT Parent Promise** – Shawna Norman
8. **Community Action Program Tulsa** – Dana James
9. **Cherokee PARENTS** – Ben King, Jennifer Kirby, Amy Thilges
10. **PCCT PAT Tulsa** - Sarah Neyman
11. **SafeCare PCCT** – Sheri Davis
12. **SafeCare NorthCare** – Dwan McDonald

University of Oklahoma HSC Analysts
for the MIECHV-3 CQI

David Bard,
Will Beasley
Geneva Marshall
Thomas Wilson
Andrew Peters
Chris Aston
Donna Wells
Som Bohora
Maleeha Shahid
Yutian Thompson

Displaying Health Data

Cases, Techniques, Solutions

Colloquium + Live-Webcast + Recording
Medical Sciences Building (MBS) 160
University of Victoria
November 28 – 30 , 1 – 3 pm PST

Please email questions to
aging@uvic.ca

University of Victoria
INSTITUTE ON AGING & LIFELONG HEALTH
island health
British Columbia OBSERVATORY Population & Public Health
IALSA
Health Sciences Center
UCF
UNIVERSITY OF CENTRAL FLORIDA

UCF

UNIVERSITY OF CENTRAL FLORIDA

The University of Oklahoma Health Sciences Center

island health

INSTITUTE ON AGING & LIFELONG HEALTH

University of Victoria

University of Victoria