

Designing, building, and deploying an interactive survey monitoring dashboard using R Shiny

Joe Murphy, Michael Duprey, Rob Chew

Government Advances in Statistical Programming Workshop
October 24-25, 2018, Bureau of Labor Statistics

Acknowledgements

- Examples for this presentation come from the National Longitudinal Study of Adolescent to Adult Health (Add Health)
- Add Health is funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development and conducted by the University of North Carolina Population Center in partnership with RTI International.
- Data presented here do not reflect actual Add Health estimates. The data were perturbed in many examples to protect the confidentiality of the study participants.

Presentation outline

- Motivation
- Static visualizations
- Interactive visualizations
- R and Shiny
- Structure (code, data, taxonomy)
- Deployment and hosting
- Benefits
- ATD Dashboard demo
- Next steps

Challenges in complex survey management

- How best to define “success” metrics
- How best to track progress toward project-defined goals
 - response rate
 - budget
 - schedule
- How to simultaneously monitor multiple (case / interviewer / data) management systems
- How to interpret reams of tabular data which may obfuscate trends or patterns

Adaptive Total Design (ATD)

- Process to identify and monitor **key features** of a survey design that are **critical to data quality**
- Similar to responsive design and adaptive design
- **Regular monitoring** of data
- Goal of ATD is to minimize **total survey error** and **costs**
- Allocates resources to maximally control errors and costs

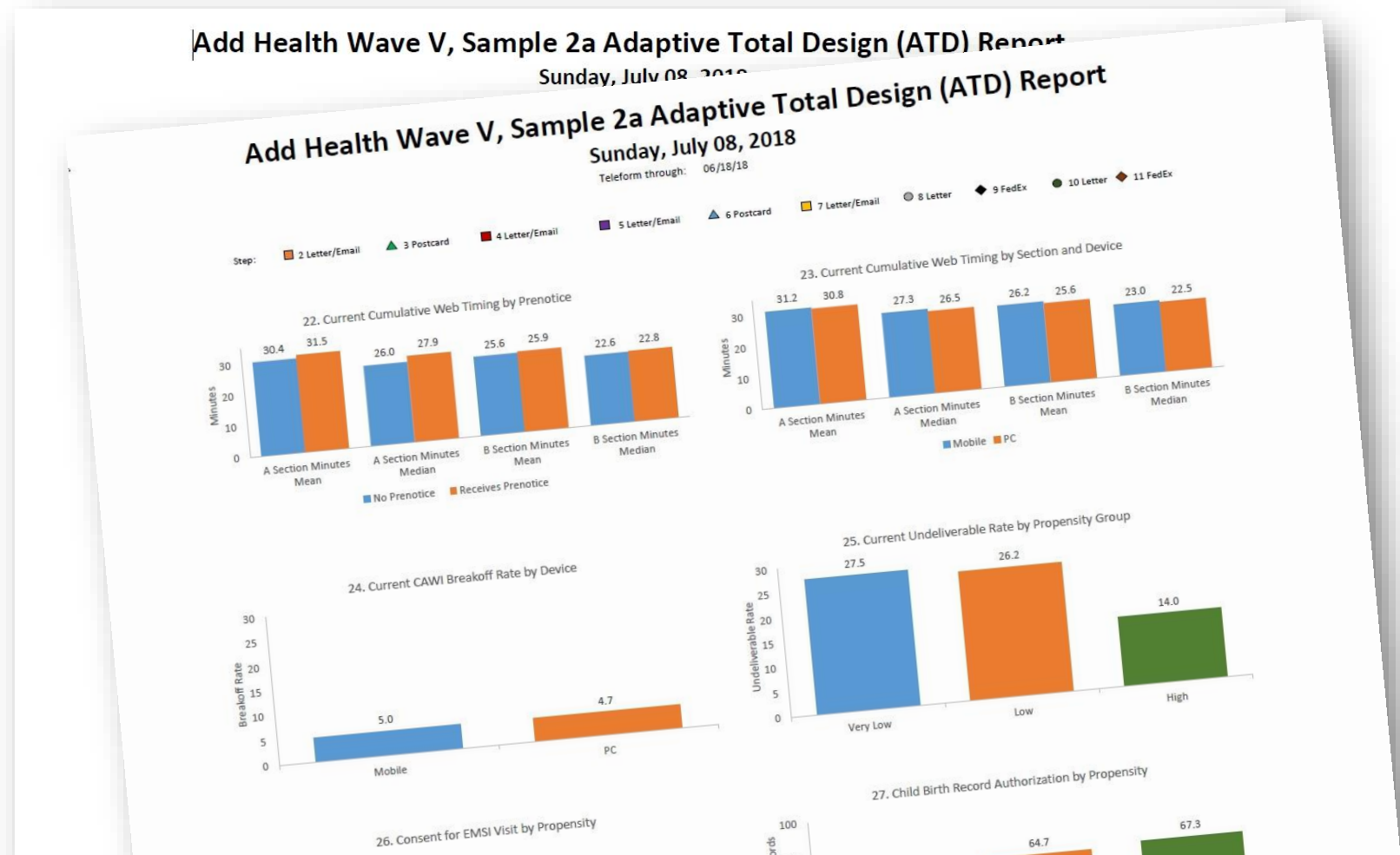
Continuous Quality Monitoring

- Start with “what metrics are key to the success of our survey?”
- Determine what other factors might affect that success
- Generate a plan for regular (real-time / hourly / daily) monitoring of data

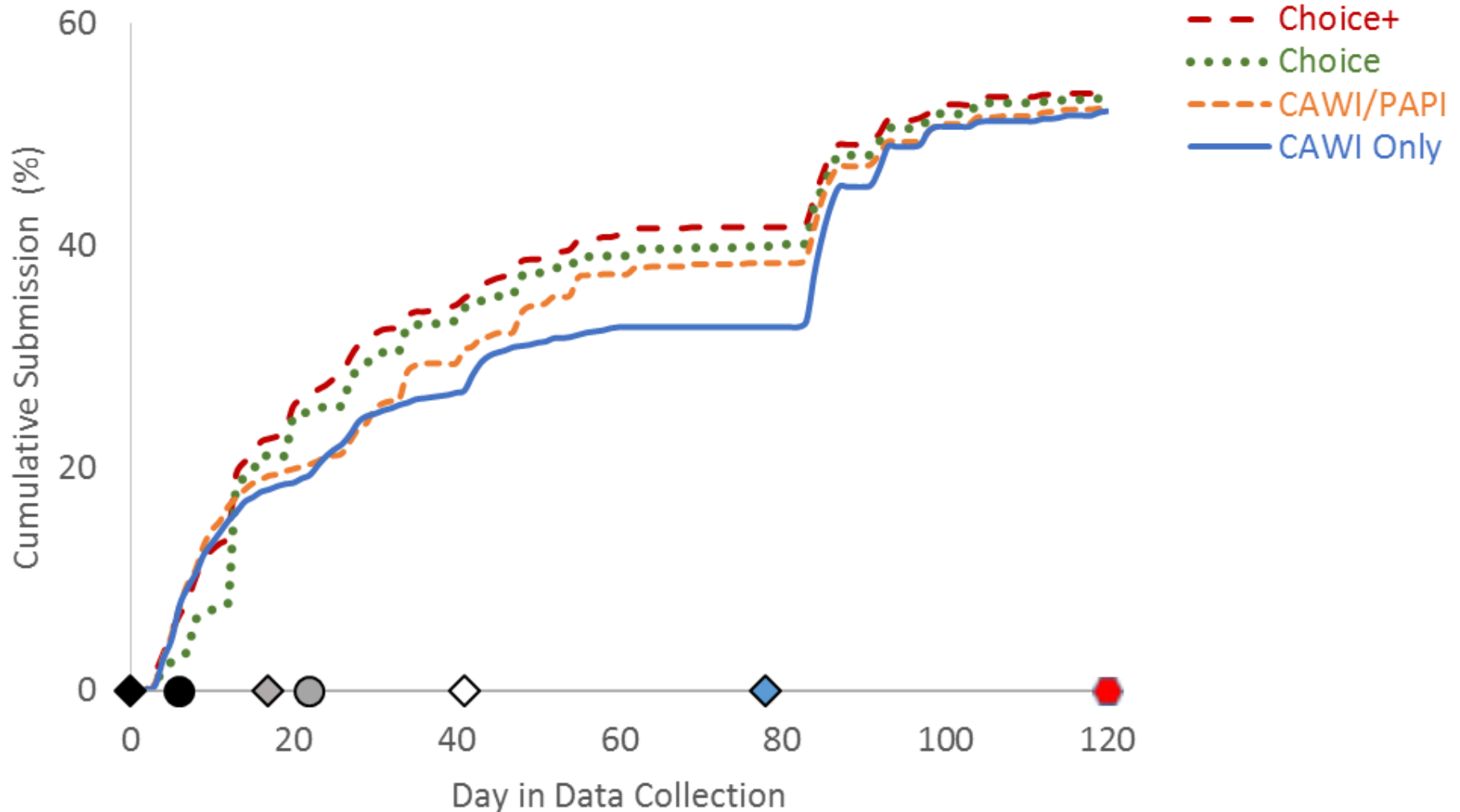
- All this is important if:
 - Your goal is to minimize total survey error and costs
 - You need to determine if / when interventions will be applied
 - You need to assess and project the outcomes of experiments during data collection

Data visualization: critical tool

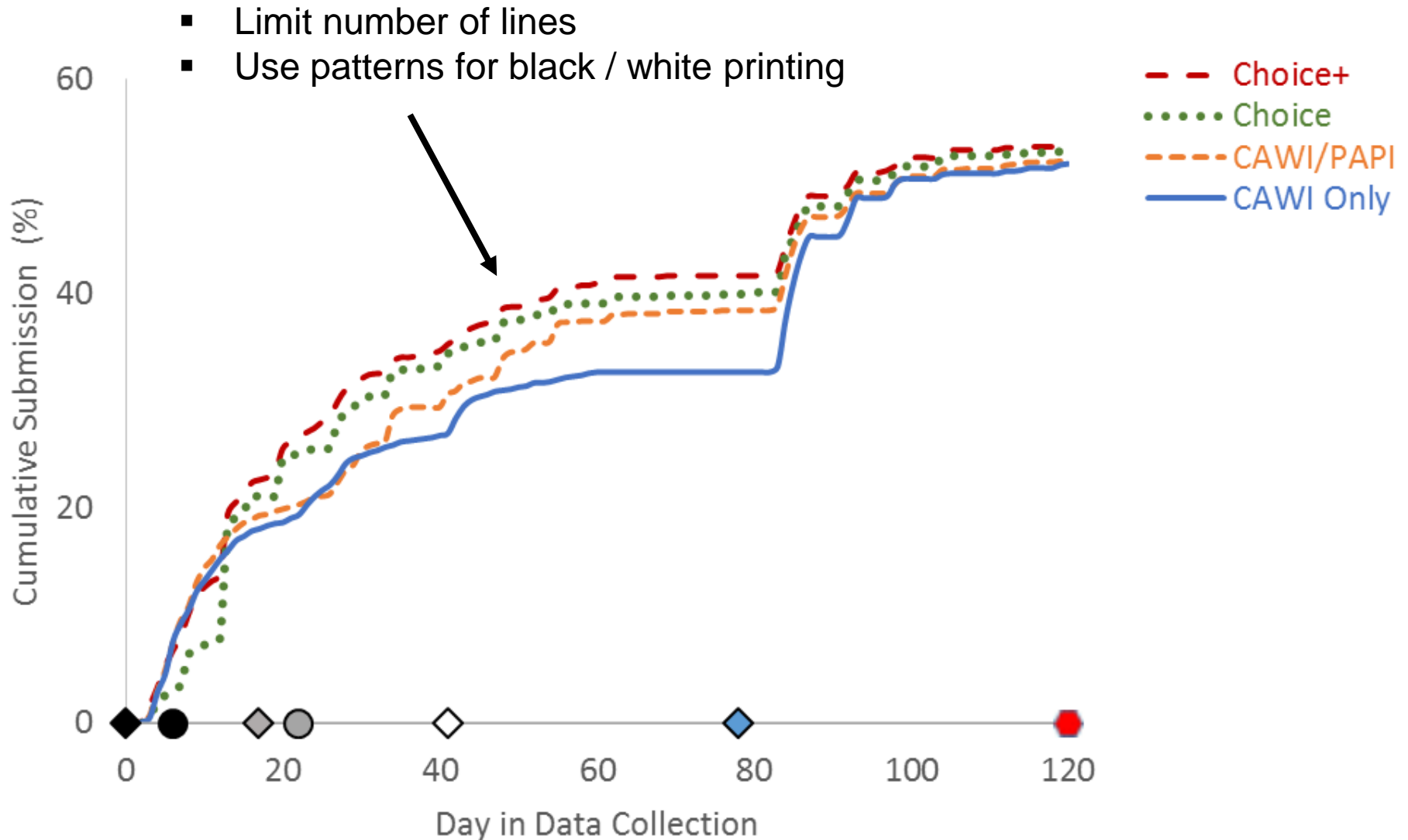
- Visualizations enable certain data to be compressed, rapidly interpreted, and easily distributed



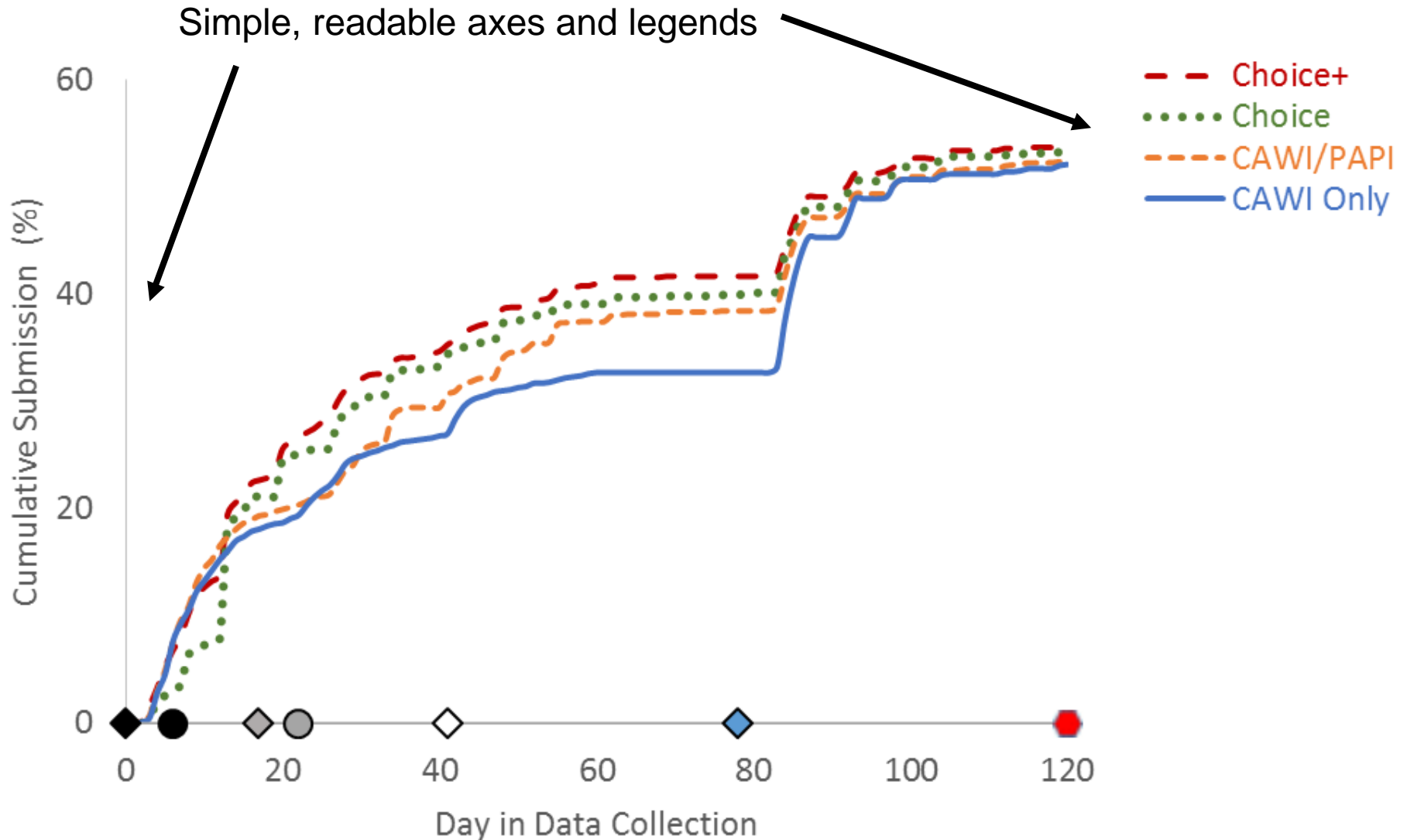
Static visualizations: response rates by mode protocol



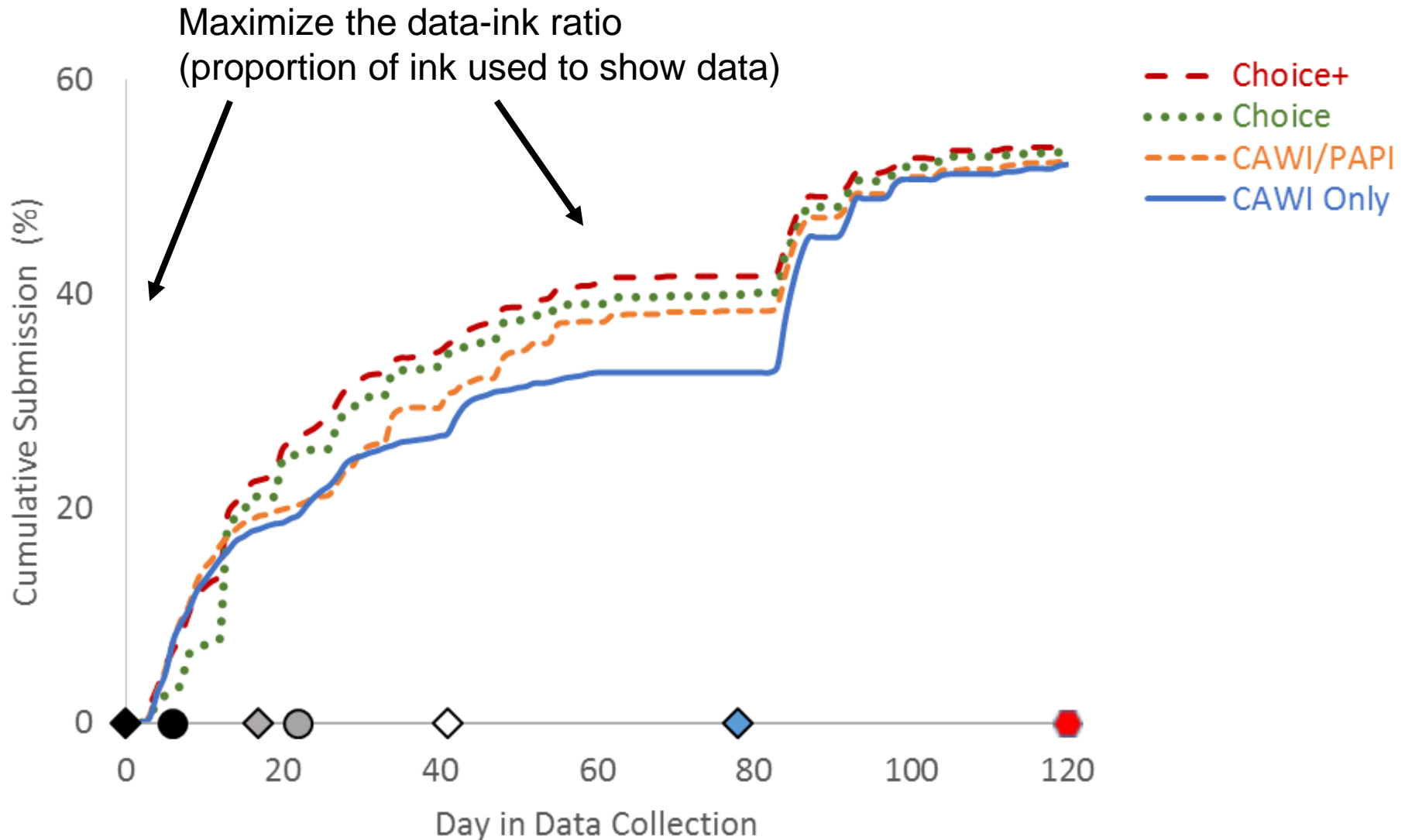
Static visualizations: response rates by mode protocol



Static visualizations: response rates by mode protocol



Static visualizations: response rates by mode protocol



Goals

- Allow users to create custom views
- Enable ability to subset data, modify display (pan, zoom, hover), and export views
- Flexibly incorporate disparate input data (e.g., survey level, respondent level, interviewer level)
- Provide web hosting for project and client teams
- Provide quick and easy setup for projects
- **Minimize user burden at every step**

Solution

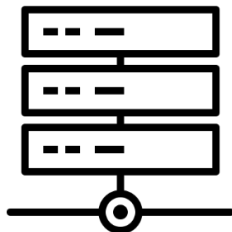
- Custom-built visualization platform programmed in **R** and **Shiny**: ATD Dashboard
- Shiny is an R package that provides a **high-level language for web application development**; avoids explicit HTML / CSS / JavaScript programming
- Can leverage robust set of packages from the R community; easily incorporate modeling
- Allows for hosting behind secure firewall with Shiny Server
- Open source and free platform (cf. Tableau)

Shiny application code architecture

- Implemented in R; however, users do not need to interface with the underlying code
- Reactive components (observe and report)

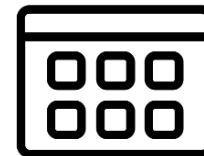
Server

- Details **logic** and **analyses** (functions)
- Generates **warnings** (passed to UI)
- Produces **plots**



User Interface (UI)

- Interacts with server file
- Details, displays, and **dynamically updates UI**, depending on a user's selections
- Displays **warnings**



Data structure

- Input data in **flat-file format** (csv file)
- Long (stacked) data
- Metadata information included in “**blocked**” sections at file head
 - Block for **critical dates** (axis markers)
 - Block for **data taxonomy**
- Format consist of **five** columns
 - Study name: **study**
 - ID: **id**
 - Variable name: **var**
 - Value: **val**
 - Date/time-stamp: **dt**

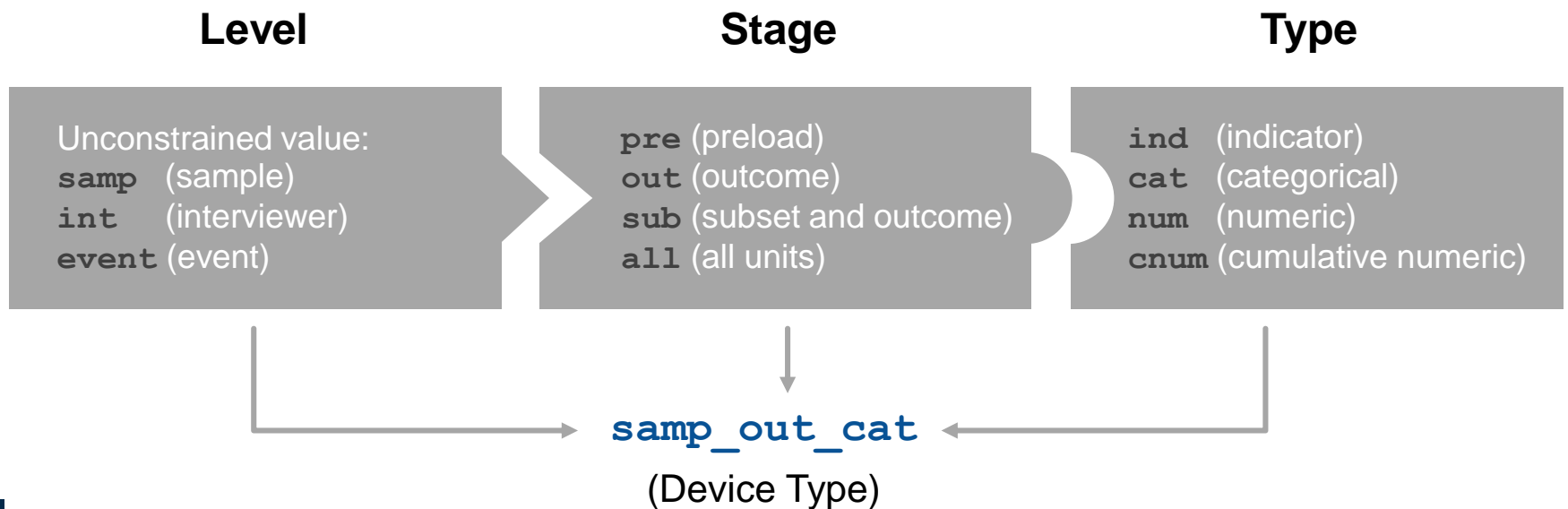
Example data

study	id	var	val	dt
...
Add Health Wave V Sample 2a	5237119	Device Type	PC	20-Mar-17
Add Health Wave V Sample 2a	5237674	Device Type	PC	17-Mar-17
Add Health Wave V Sample 2a	5237765	Device Type	Mobile	17-Mar-17
Add Health Wave V Sample 2a	5237981	Device Type	Mobile	27-Mar-17
Add Health Wave V Sample 2a	5238472	Device Type	Mobile	15-Mar-17
...

Data taxonomy

- Provides **compact attribute vocabulary**, easy to input into flat file
- Provides mechanism to limit UI selection options to **only logical combinations** and invoke **variable mathematical operations**
- **Mapped** to data in csv file

Type-class system



ATD Dashboard

Data

study	id	var	val	dt
...
Add Health Wave V Sample 2a	5237119	Device Type	PC	20-Mar-17
Add Health Wave V Sample 2a	5237674	Device Type	PC	17-Mar-17
Add Health Wave V Sample 2a	5237765	Device Type	Mobile	17-Mar-17
Add Health Wave V Sample 2a	5237981	Device Type	Mobile	27-Mar-17
Add Health Wave V Sample 2a	5238472	Device Type	Mobile	15-Mar-17
...

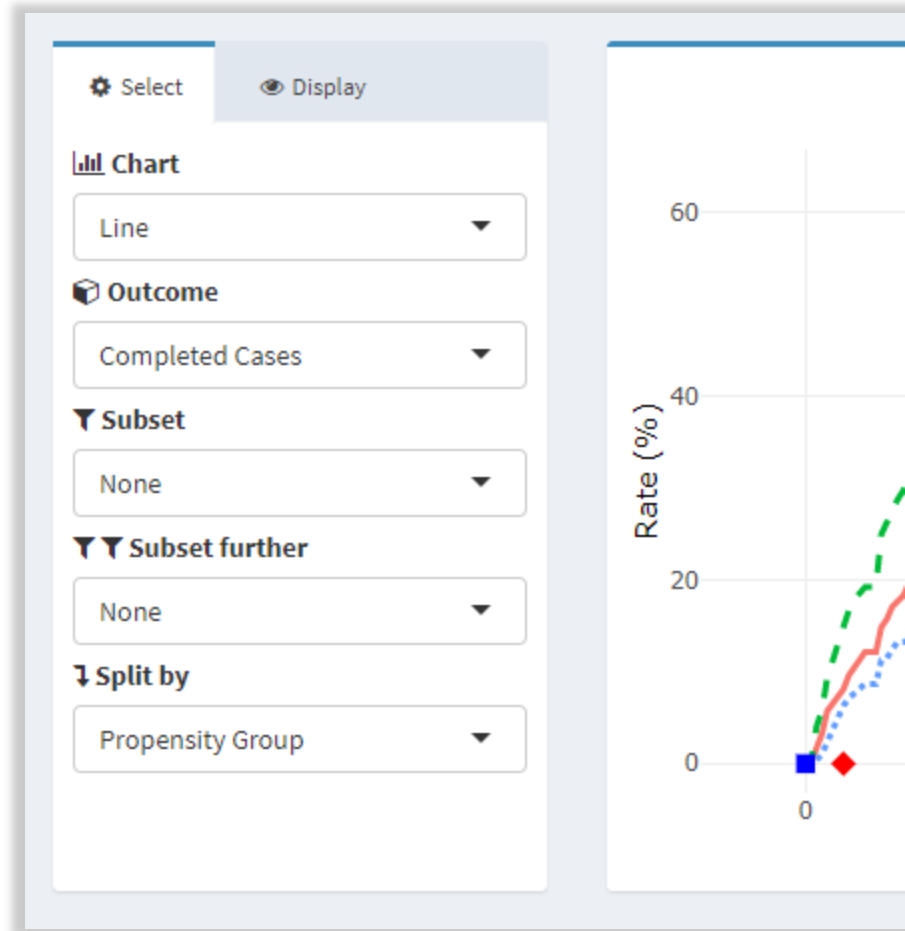
Map

study	id	var	val	dt
Add Health Wave V Sample 2a		C:Undeliverables	samp_out_ind	
Add Health Wave V Sample 2a		C:Completion Mode	samp_out_cat	
Add Health Wave V Sample 2a		C:Device Type	samp_out_cat	
Add Health Wave V Sample 2a		C:Self R. Health: Excellent	samp_out_cat	
Add Health Wave V Sample 2a		C:Self R. Health: Fair	samp_out_cat	
Add Health Wave V Sample 2a		C:Undeliverables	samp_out_ind	
...

Data taxonomy

UI drop-down options reactively updated to allow logical combinations, conditioned on prior selections

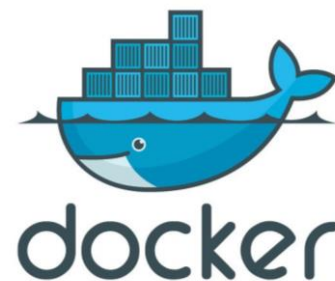
- Choose **plot type** →
- Select primary **outcome of interest** →
`samp_out_cat`
- Filter to **subset** →
(e.g., `samp_sub_ind`)
- Filter to (second) **subset** →
- Separate** outcome into categories →
`samp_out_cat`



Deployment and hosting

Shiny Server

- Apps deployed with RStudio Shiny Server, hosted on local or remote computers
- Can be deployed to the web or shared within a closed network behind a firewall with authentication for site access
- For reproducibility and to aid in future deployment to other servers, both R Shiny Server and an NGINX web server (used as a reverse proxy to redirect requests) were configured and deployed in Docker containers



ATD Dashboard Examples

Select Display

Chart

Line

Outcome

Completed Cases

Subset

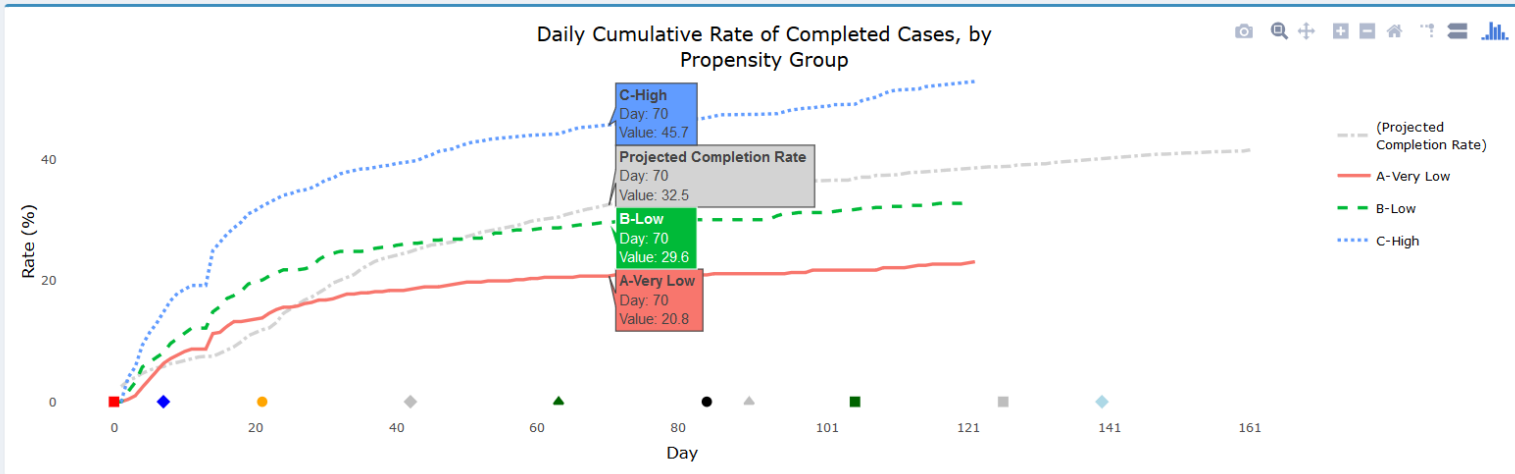
None

Subset further

None

Split by

Propensity Group



Select Display

Chart

Bar

Outcome

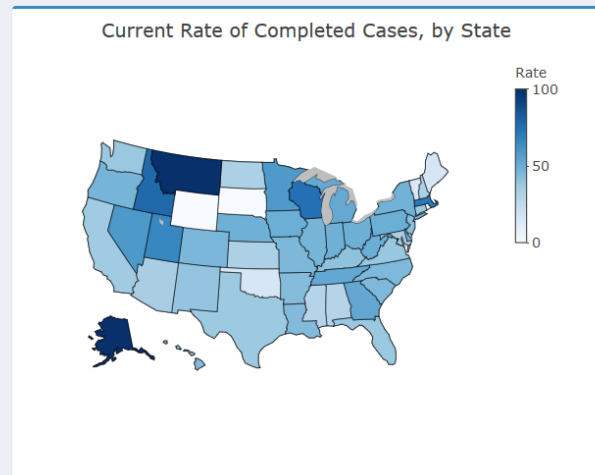
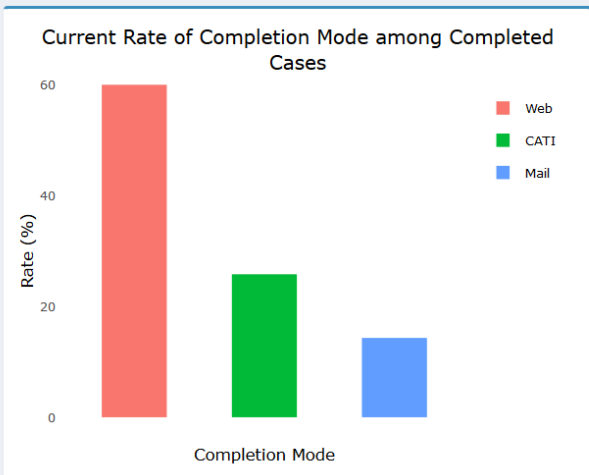
Completion Mode

Subset

Completed Cases

Subset further

None



Select Display

Chart

Map

Outcome

Completed Cases

Subset

None

Subset further

None

Select ⚙

Display 👁

Chart

Line

Outcome

Completed Cases

Subset

None

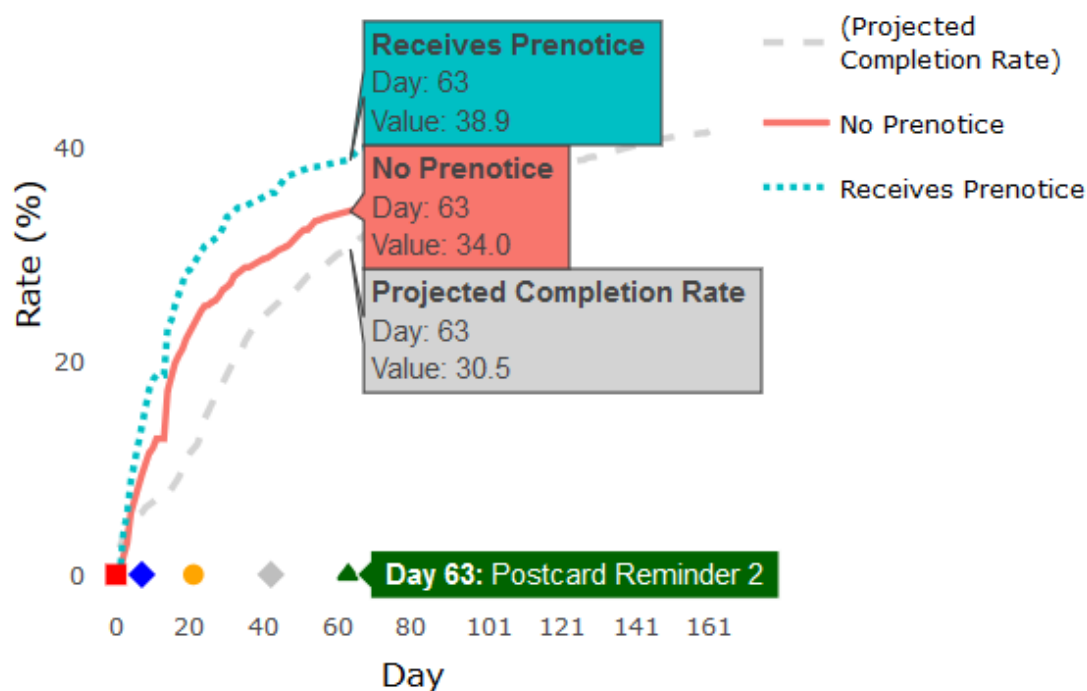
Subset further

None

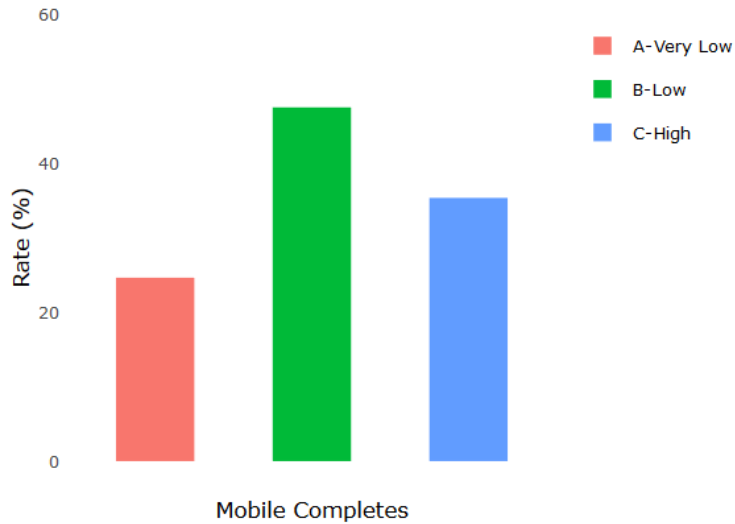
Split by

Prenotice

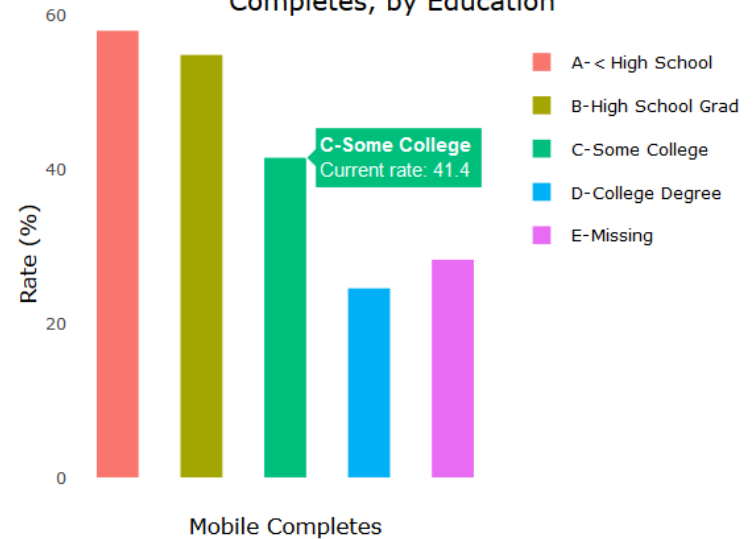
Daily Cumulative Rate of Completed Cases, by Prenotice



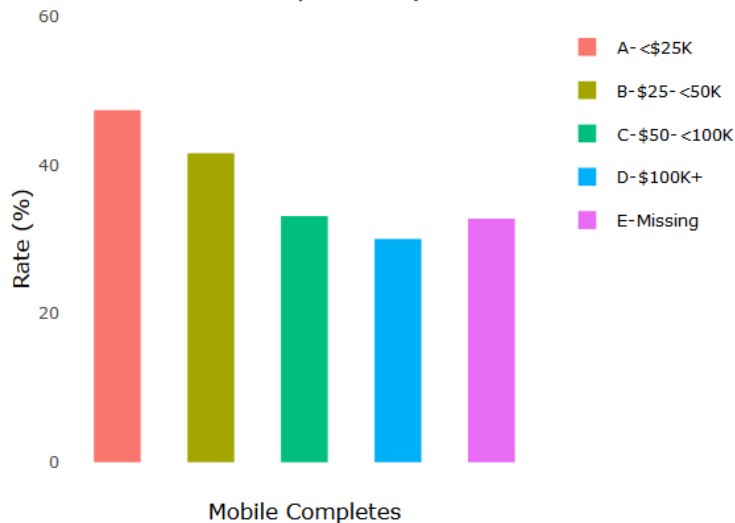
Current Rate of Mobile Completes among Web Completes, by Propensity Group



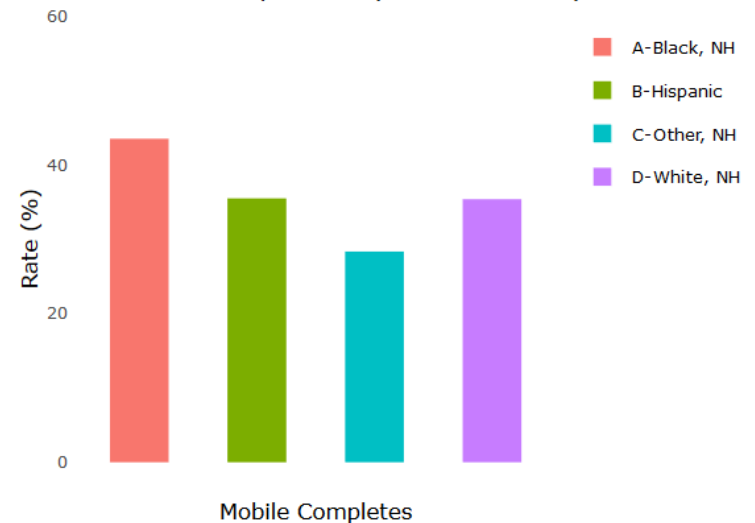
Current Rate of Mobile Completes among Web Completes, by Education



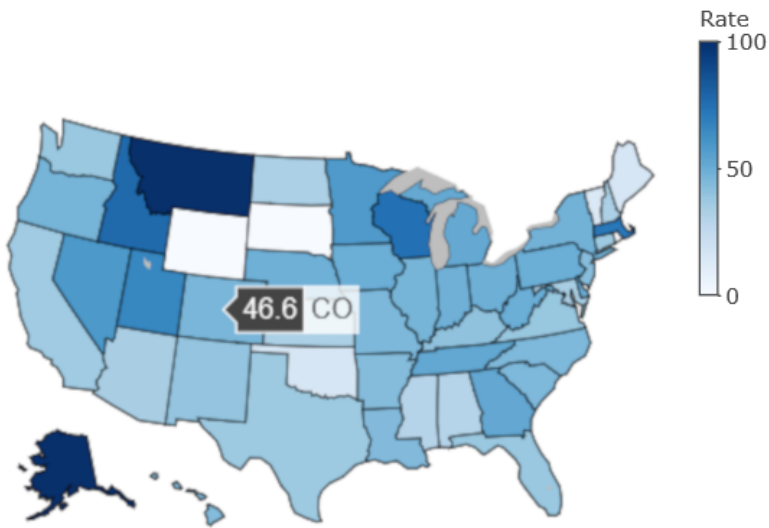
Current Rate of Mobile Completes among Web Completes, by Income



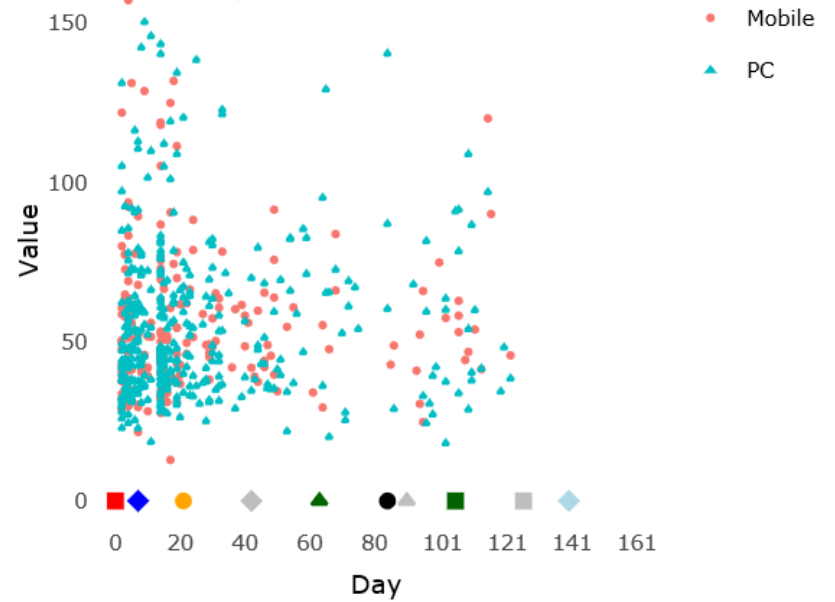
Current Rate of Mobile Completes among Web Completes, by Race/Ethnicity



Current Rate of Completed Cases, by State



Daily Interview Minutes: Total among Web Completes, by Device Type



Reception

- **Internal** users: continual monitoring of data collection helpful for discussions related to production goals and strategies within the current sample (i.e. use of and timing of additional mailings)
- **External** users: format of “everything on a page” found to be useful to get up to speed with the current results at a glance and be prepared to discuss implications for the design of future samples with the internal team

Impact

- At the beginning of a study, it can be a challenge to define the key metrics for data collection
- More flexibility in monitoring and planning for future sample releases
- Many potential metrics and narrowing it down can be difficult
- Having the dashboard and needing to decide what to monitor spurred conversations about goals that might not otherwise have taken place

Summary of benefits

- Ability to monitor survey in near real time for decision making
- Only most important information (not reams of tables) displayed
- Leverages best practices in graphic design
- Non-programmer team members can gain insights from data; clients can interact with data directly
- No licenses required for individuals
- Codebase is reusable, easily transportable
- Codebase is easily extensible, thanks to R
- Dashboards can be shared across organization, avoiding costly licensing fees

Next steps

- Incorporate weighted data and statistical tests
- Pull-up data tables
- Allowing multiple layers of graphs

Thank you!

Contact

Joe Murphy

jmurphy@rti.org

RTI
INTERNATIONAL