Video as data & documentation

Rick O. Gilmore

2019-09-18 11:56:41

Preliminaries

I thank NSF, NIH, the Alfred P. Sloan Foundation, SRCD, the Lego Foundation, and the James S. McDonnell Foundation for support.

Overview

- Video as data
- Video as documentation
- Sharing ethically & securely with Databrary
- Accelerating discovery

I'll be talking about video as data and documentation, how video can be shared ethically and securely, and how video sharing will accelerate discovery in the psychological sciences.

Video as data & documentation

Adolph, Gilmore, & Kennedy, 2017

Video as data

Frank, 2014

Well, let me explain, or better, show. Tell me what's going on in this short clip?

Frank, 2014

And how about this one?

That's right. It's a mental abacus.

```
## PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, pleas
```

Frank, 2014

And the video let you see the phenomenon directly BEFORE you even read the paper.

DeLoache, Uttal, & Rosengren 2004

Another example comes from the work of Judy DeLoache and colleagues on some children's surprising errors of spatial scale. DeLoache and colleagues observed that children sometimes tried to sit on tiny doll-sized chairs or insert themselves into tiny doll-sized cars after playing and interacting with full-sized chairs or toys.

When they submitted the paper to Science, the editors didn't believe them.

If a picture is worth 1,000 words...

A video is worth...

DeLoache, 2014

DeLoache 2014

. . . a paper in *Science*

Video. . .

- Captures (& preserves)
- Shows (& helps tell. . .)
- Expands the scope of inquiry
- Provides unparalleled opportunities for reuse

Video captures and preserves aspects of behavior that other measures often neglect; it shows what participants do and helps tell the story; it expands the scope of inquiry to include other behaviors; and it provides unparalleled opportunities for reuse.

Video also has unique strengths as research documentation.

Video as documentation

Baker, 2016

These are the results.

Baker, 2016

Baker, 2016

Gilmore & Adolph, 2017

Karen Adolph and I argue that video recordings of research procedures can make behavioral science more reproducible. Here are some examples.

Yu 2016

Let's say you're interested in where children and their parents look to explore the role of joint attention in early word learning.

This video from Databrary shows what the raw video from eyetracking systems used by Yu et al. actually looks like.

Even audio

Audio recordings can also be shared.

Cole, Gilmore, Perez-Edgar, & Scherf 2017

I'm largely a vision scientist, but more generally I'm interested in perceptual and motor development. In collaboration with other colleagues at Penn State, we've been exploring behavioral and neural responses to maternal speech in different affective prosodies.

Cole, Gilmore, Perez-Edgar, & Scherf 2017

Here are two examples. The full set of recordings are shared on Databrary for the research community to use.

Shared video & audio recordings

- Make research more transparent
- Bolster the reproducibility of procedures
- Accelerate the adoption of new research techniques

- Strengthen findings
- Maximize the return on public investments in research

Sharing video ethically & openly

Video must be protected

Sharing video ethically requires protections.

- Faces, voices
- Home interiors
- Behaviors that may embarrass participants

Video is potentially identifiable. It contains faces and voices. It might contain images of home interiors or capture behaviors that could embarrass participants.

Open sharing advances discovery

At the same time, open sharing, with limited restrictions on appropriate uses, promises to advance discovery.

How do we strike a balance between the two?

- Citation expected, but
- No requirement for co-authorship
- No pre-approval of research questions

Databrary does not pre-approve research questions nor enforce requirements for co-authorship, but we expect that any researcher using our materials to cite the source.

Requires. . .

- researchers secure participant permission to share
- IRB/ethics board approval to share
- institutional approval for access

Moreover, Databrary requires researchers to secure permission from research participants in order to share their video or audio data. Databrary requires researchers seek ethics board approvals as required by a researcher's institution. And most importantly, Databrary restricts access to researchers who have been authorized by their institutions to access the site.

Permission to share

- Standard language via templates
- When to ask & how decided by local ethics boards/IRBs
- Sample scripts and videos

Databrary has developed standard language for seeking permission to share and provides guidance about when and how to seek ethics board approval. We share sample scripts for talking with participants alongside video exemplars.

Some researchers might be uncertain about how to ask parents or participants for permission to share video or audio, so Databrary includes videos that show how easy it is.

Restricting access

- Researchers at institutions
- Formalized by an institutional agreement
- Protects participants, researchers, & institutions

Only researchers whose institutions have granted explicit access are permitted to share data. Student or staff affiliates an authorized researcher supervises may download or resuse Databrary materials if granted access by their supervisor. A researcher's institution must sign a formal institutional agreement with NYU, Databrary's host institution, before a researcher gains full access.

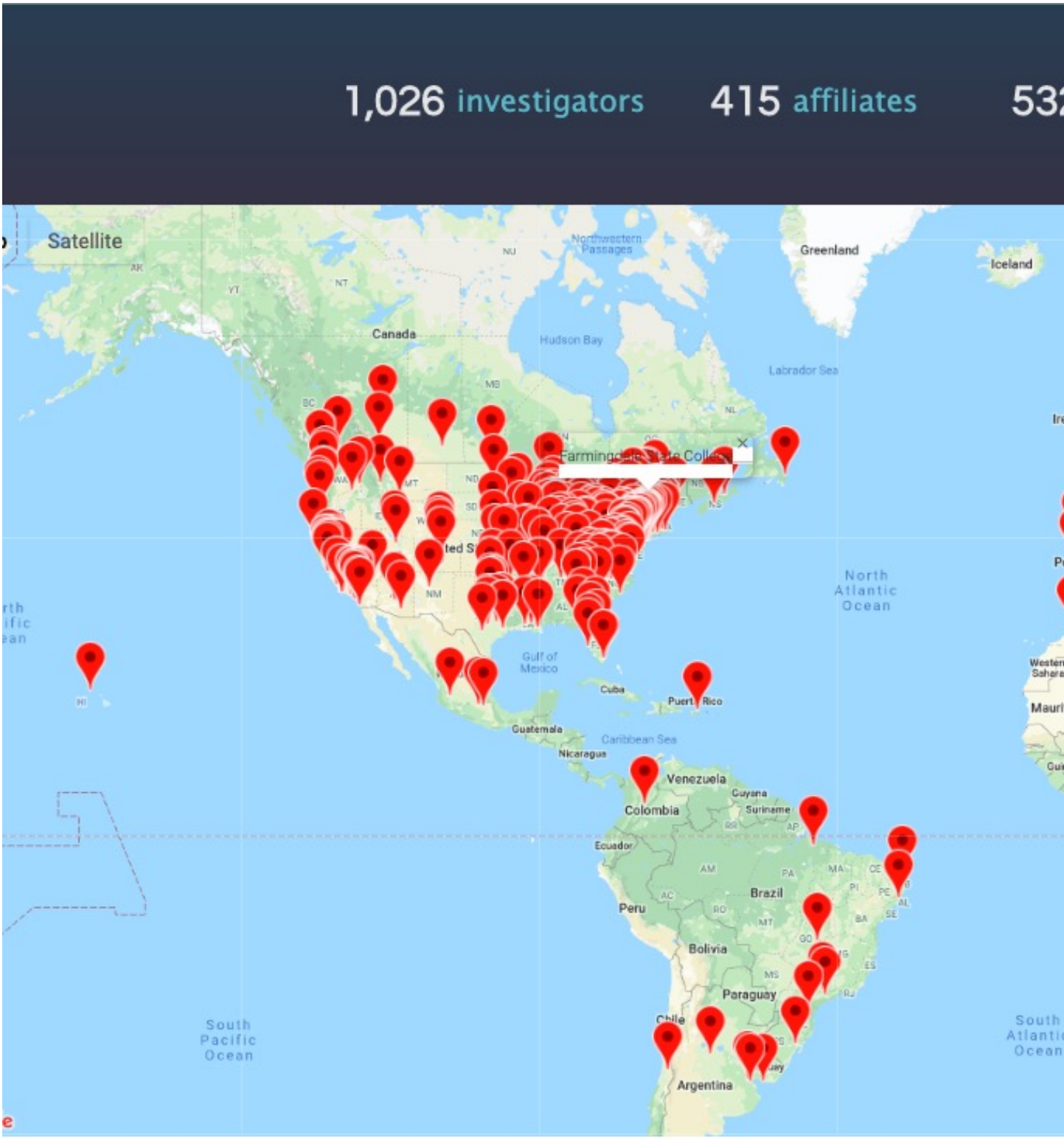
This policy framework protects participants, researchers, and institutions.

Standardized levels of sharing

Databrary has standardized the levels of sharing into categories; every file on the site is tagged with the level of sharing permitted.

Symbol	Release Level	Defi
?	Unknown	Consent missing or not obtain and is restricted to collaborat
	Private	This content is not shared and
	Authorized Users	This content is restricted to au may not be redistributed in an
	Learning Audiences	This content is restricted to au who may use clips or images f informational or educational p may be videotaped or recorde recordings may then be made internet (e.g., YouTube).
	Public	This content is available openl

Private files are restricted to the authorized researcher who uploaded them and any other parties they grant access to.



As this map indicates, the Databrary policy framework has been agreed to by hundreds of institutions across the globe.

Accelerating discovery

Need

- Annotation tools
- More (bigger, denser, diverse) data
- Platforms for discovery

To exploit the power of video to accelerate discovery in behavioral science, we need annotation tools, bigger data, and software tools that go beyond passive data storage but are truly platforms for discovery.

Databrary is one such tool.

Datavyu, a free open-source, annotation tool is another.

Adolph, Tamis-LeMonda, & Gilmore 2017

We're also creating a large-scale, diverse and rich data set for others to exploit via the Play & Learning Across a Year (PLAY) Project.

- $n = 900$ infant/mother dyads, $n = 300$ 12-, 18-, and 24-month-olds
- Demographics, health status, media use, temperament
- Modified MB-CDI (interview, video-recorded, English/Spanish item-level equivalents)
- Open protocol, video documentation

The entire PLAY project protocol, including videos of procedures, coding definitions, survey items, etc. is openly shared.

Repositories as Platforms for discovery

Platforms for discovery require software that interacts with stored and shared data in interesting and useful ways.

We're making progress on this front, too.

Tamis-LeMonda, 2013

Here's one of the largest currently shared datasets on Databrary from Catherine Tamis-LeMonda.

Maybe you're interested in seeing whether it is large and diverse enough for your needs.

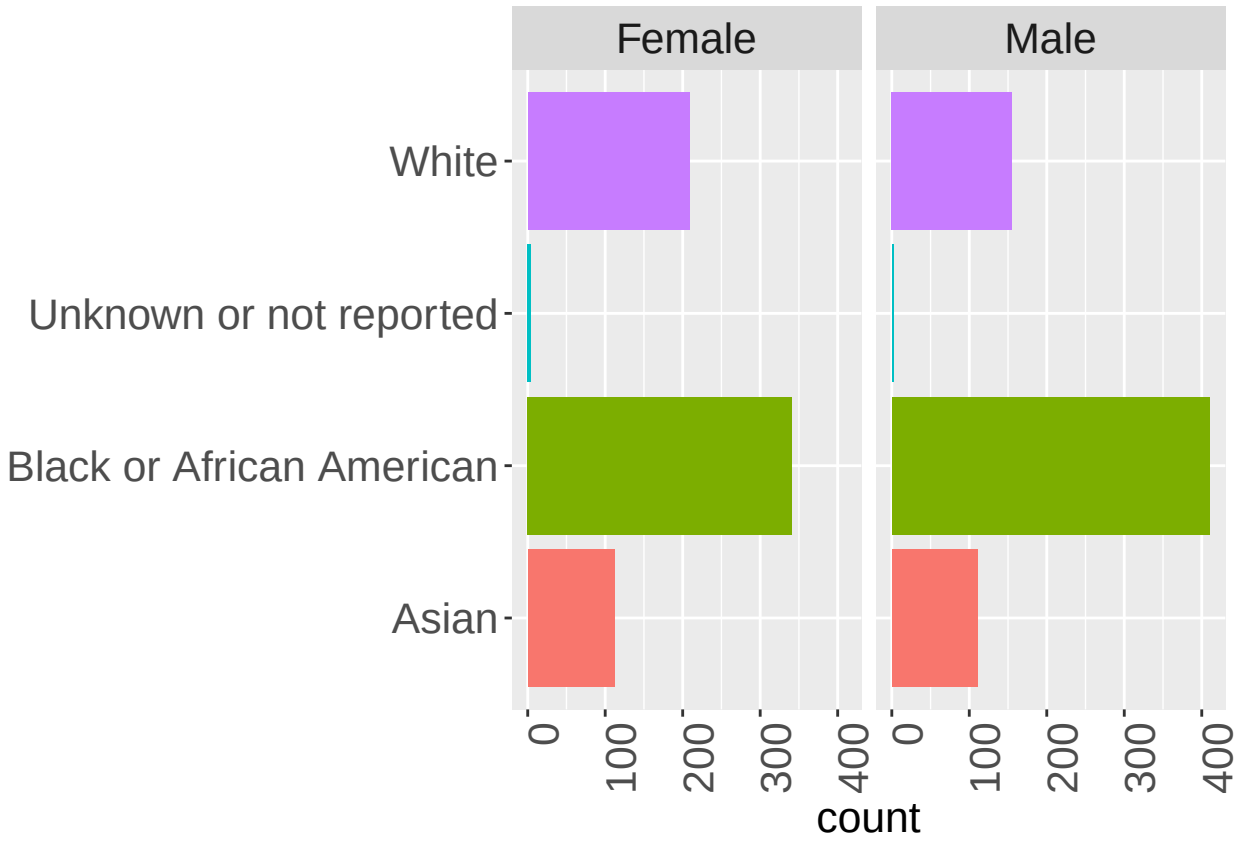
```
vol_8 <- databraryapi::download_session_csv(vol_id = 8)
xtabs(formula = ~ participant.gender + participant.race, data = vol_8) %>%
  knitr::kable(.)
```

		Asian	Black or African American	Unknown or not reported	White
	7	0	0	0	0
Female	0	112	341	4	209
Male	0	111	410	2	155

Using a few commands in the R programming language, you can securely download the entire demographic dataset from Databrary and plot a tabular summary.

```
race_gender <- vol_8 %>%
  filter(participant.gender %in% c('Male', 'Female')) %>%
  ggplot() +
  aes(x = participant.race, fill = participant.race) +
  facet_grid(. ~ participant.gender) +
  geom_bar(stat="count") +
  coord_flip() +
  theme(axis.text.x = element_text(angle = 90,
                                   hjust = 1, size = 16),
        axis.text.y = element_text(size = 16),
        axis.title.y = element_blank(),
        axis.title = element_text(size = 16),
        strip.text = element_text(size = 16),
        legend.position = "none")
```

Or you can create a plot of the data for a report to your collaborators.



```
databraryapi::login_db(email = params$db_login)
```

```
## [1] TRUE
```

```
play_session_id <- 26295
play_vol_id <- 444

vol_444_assets <- list_assets_in_session(vol_id = play_vol_id,
                                         session_id = play_session_id) %>%
  select(name, asset_type, asset_id)
```

Maybe you're interested in reusing the PLAY project pilot data and want to know what other types of data files are available.

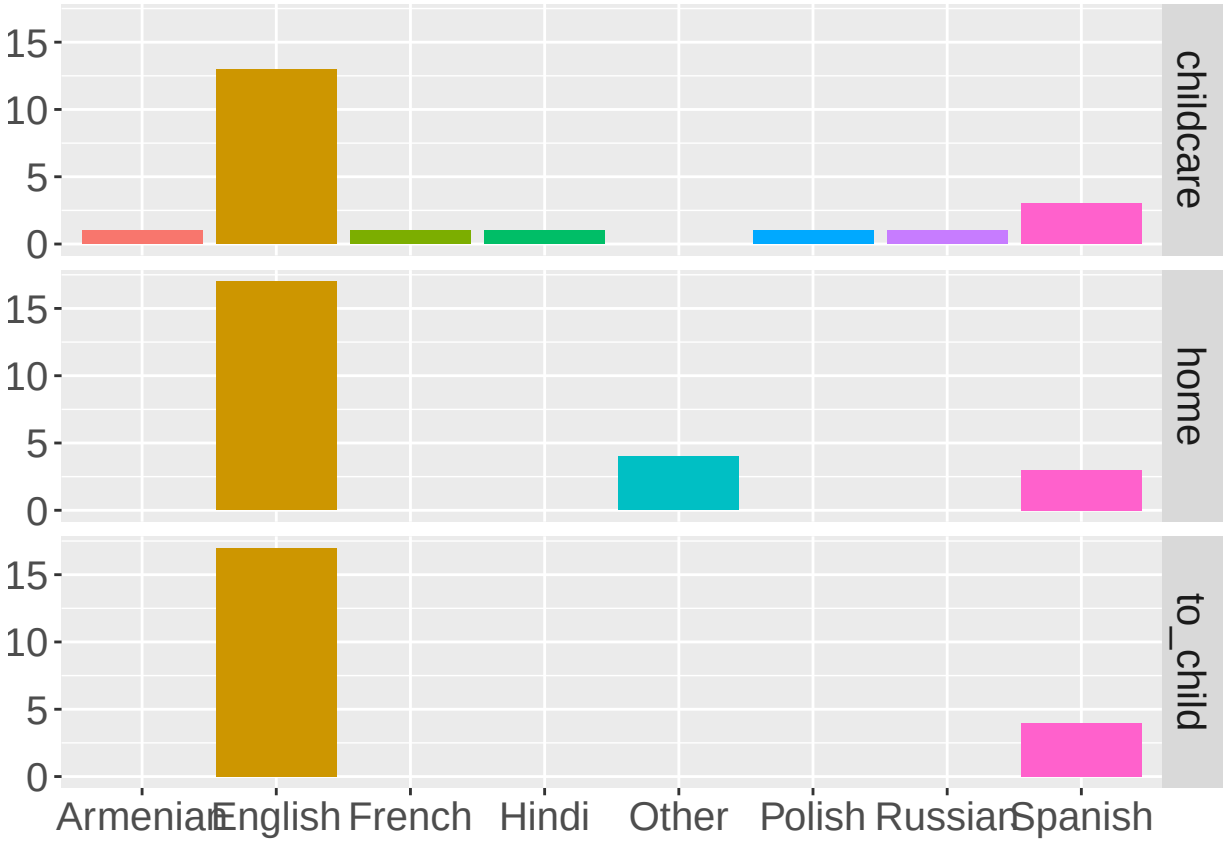
---

name	asset_type	asset_id
childcare	Comma-separated values	159177
locomotion	Comma-separated values	117092
child-birth	Comma-separated values	116791
family	Comma-separated values	116790
sleep	Comma-separated values	116789
language-exposure	Comma-separated values	116787

```
lang_exp <- databraryapi::read_csv_data_as_df(session_id = play_session_id,
                                              asset_id = 116787)

lang_exp_plot <- lang_exp %>%
  ggplot(.) +
  aes(x = language, fill = language) +
  facet_grid(exposure_context ~ .) +
  geom_histogram(stat='count') +
  theme(axis.text = element_text(size = 15),
        strip.text = element_text(size = 15),
        axis.title = element_blank(),
        legend.position = "none")
```

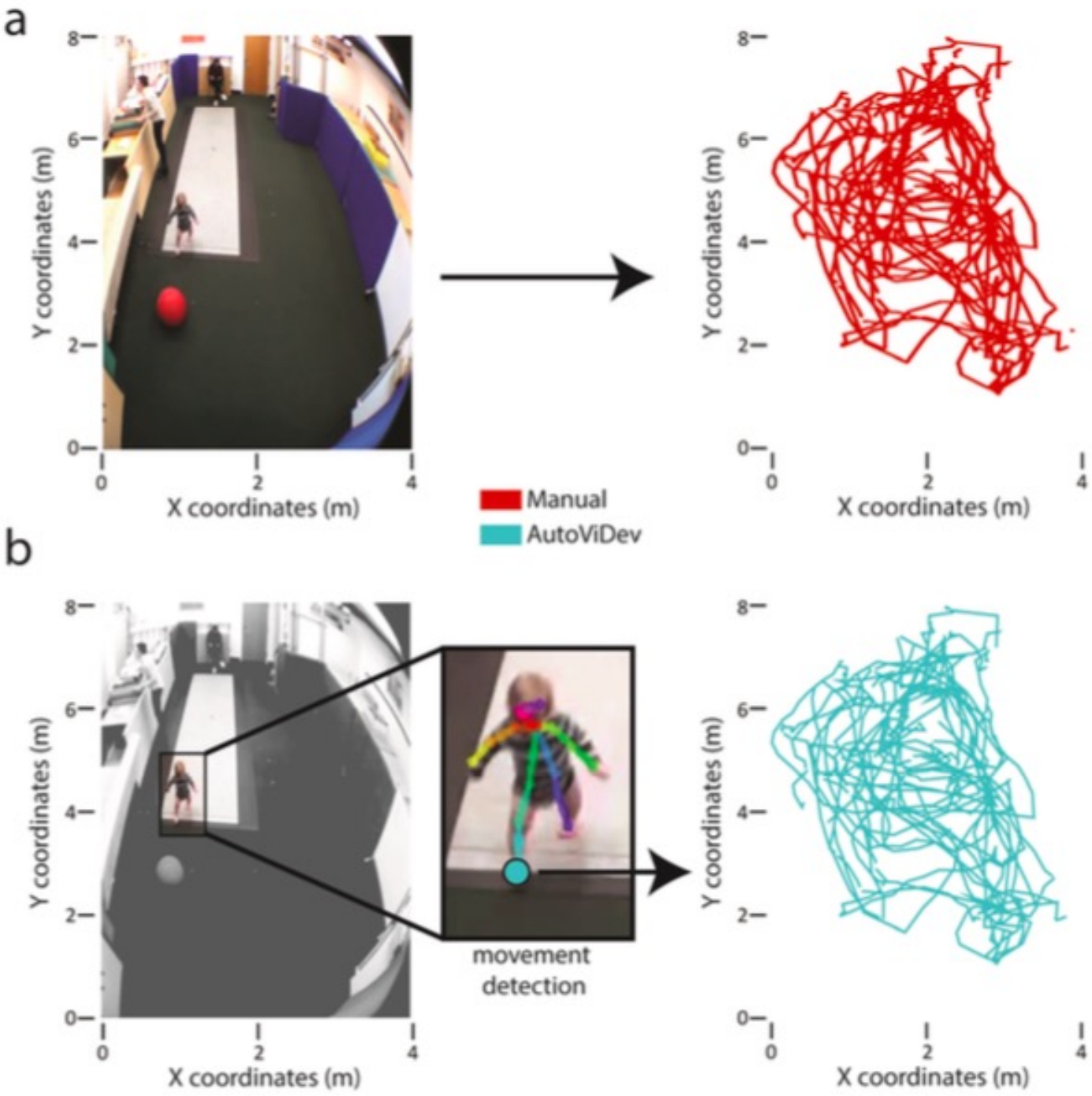
Since you're a language researcher, you're especially interested in the types of languages children are exposed to and in what settings.



A few commands later, those data are available to you.

Your browser does not support the video tag.

Video is just one type of time series



Ossmy, Gilmore, & Adolph, 2019

Next steps...

Busting some myths

- Data X can't be shared...
- Open sharing means public
- My data are mine

Each one, reach one

> "*SRCD regards scientific integrity, transparency, and openness as essential for the conduct of research and its application to practice and policy.*"

Society for Research in Child Development (SRCD) Policy

> "*These values apply to research conduct, to the teaching of scientific methods, and to the translation of science into practice and policy.*"

SRCD Policy

> "*SRCD regards it as an essential responsibility of the producers and consumers of child development research to act ethically, conscientiously, and with integrity...*"

> "*...and to strive for transparency and openness in research practice, publication, and dissemination.*"

SRCD Policy

Investing in the future

- ...of robust & reproducible social science

A modest proposition: The 15% solution

Federal funding agencies should set aside...

- 5% of the research budget for data *repository operational support*
- 5% of the research budget for *data curation* activities
- 5% of the research budget for *secondary analyses* of shared data

Thank you

---

rogilmore@psu.edu https://gilmore-lab.github.io https://gilmore-lab.github.io/2019-08-28-apsa/ @rogilmore

Materials

This talk was produced on 2019-09-18 in RStudio version 1.2.1335 using R Markdown and the reveal.JS framework. The code and materials used to generate the slides may be found at https://github.com/gilmore-lab/2019-08-28-apsa/. Information about the R Session is as follows:

```
## R version 3.5.3 (2019-03-11)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Mojave 10.14.6
##
## Matrix products: default
## BLAS: /System/Library/Frameworks/Accelerate.framework/Versions/A/Frameworks/vecLib.framework/Versions
## LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] qrcode_0.1.1            forcats_0.4.0
##  [3] stringr_1.4.0           dplyr_0.8.1
##  [5] purrr_0.3.2             readr_1.3.1
##  [7] tidyr_0.8.3             tibble_2.1.3
##  [9] ggplot2_3.2.0           tidyverse_1.2.1
## [11] databraryapi_0.1.6.9001
##
## loaded via a namespace (and not attached):
##  [1] revealjs_0.9      tidyselect_0.2.5  xfun_0.8
##  [4] reshape2_1.4.3    haven_2.1.0       lattice_0.20-38
##  [7] colorspace_1.4-1  generics_0.0.2    htmltools_0.3.6
## [10] yaml_2.2.0        rlang_0.4.0       R.oo_1.22.0
## [13] pillar_1.4.1      glue_1.3.1        withr_2.1.2
## [16] R.utils_2.9.0     modelr_0.1.4      readxl_1.3.1
## [19] plyr_1.8.4        munsell_0.5.0     gtable_0.3.0
## [22] cellranger_1.1.0  R.methodsS3_1.7.1 rvest_0.3.4
## [25] codetools_0.2-16  evaluate_0.14     labeling_0.3
## [28] knitr_1.23        curl_3.3          highr_0.8
## [31] broom_0.5.2       Rcpp_1.0.1        scales_1.0.0
## [34] backports_1.1.4   webshot_0.5.1     jsonlite_1.6
```

```
## [37] hms_0.4.2         packrat_0.5.0     digest_0.6.19
## [40] stringi_1.4.3     keyring_1.1.0     grid_3.5.3
## [43] cli_1.1.0         tools_3.5.3       magrittr_1.5
## [46] lazyeval_0.2.2    crayon_1.3.4      pkgconfig_2.0.2
## [49] xml2_1.2.0        lubridate_1.7.4   assertthat_0.2.1
## [52] rmarkdown_1.13    httr_1.4.0        rstudioapi_0.10
## [55] R6_2.4.0          nlme_3.1-140      compiler_3.5.3
```