

CENTER ON RURAL INNOVATION

WORKING WITH PRIMARY DATA

`cori.data.fcc::` an example of data as code

Olivier Leroy

Center On Rural
Innovation

2024-10-20

SLIDES ARE ONLINE:

- Realized in [Quarto](#) (Reavealjs)
- Running R and cori.data.fcc
- Hosted/deployed in [GitHub](#)
https://ruralinnovation.github.io/prez_aube



GOALS:

1. Primary data is great! (but can be
2. Use code (other pain?) to work with data
3. Packaging it will make life easier

PLAN: RETRACE OUR JOURNE

[TODO: rework at the end]

HI, I AM OLIVIER SENIOR DATA ENGINEER

- Working with primary data: “messy, unstructured data” upon us

Analytics And Data Science

Data Scientist: The Sexiest Job of the 21st Century

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)

Source: [Harvard Business Review](#)

- Socials: [LinkedIn](#) | [Mastodon](#) | [Personal Website](#)

DATA AT CENTER ON RURAL I

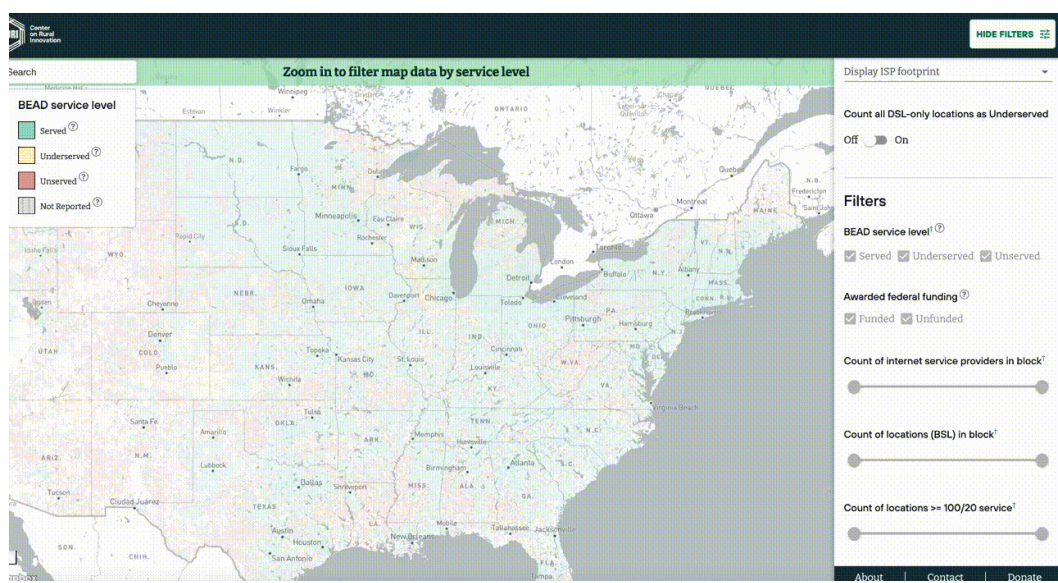
(CONT)

1. Rural data: we can't work with aggregate
deeper

I plan to delete this unless good points again

**BROADBAND: AN IMPORTANT
PART OF OUR WORKS**

PUBLIC-FACING APPS:



Rural Broadband Mapping Tool

Explor


Click on a tract

- ☒ Hazard risk
- ☐ Very Low



Bro Ris


RESEARCH WORK: FROM IDEA REPORT





Amanda Weinstein

May 3rd at 9:39 AM

Is it possible to give me the histogram of locations for Calix frns?

 1





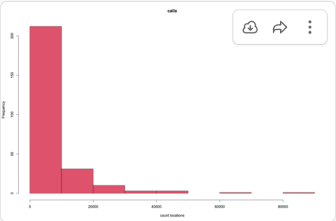
Olivier Leroy

May 3rd at 10:05 AM

@Amanda Weinstein

 this was some "dirty work" (keep in mind we just have around 266 calix ISP identified with an FRN);

image.png



Min. 1st Qu. Median Mean 3rd Qu.

2 1229 3522 6660 8100

Max.

83057

For reference the company that is "serving" 2 locations is "choctaw telephone company"



Amanda Weinstein

May 3rd at 10:12 AM

PERFECT! I want to check to see if our definition of small could eliminate some Calix customers

Beyond
Connections

The Role of
Broadband
Economic
Resiliency

Dr. Amanda Weinstein

Director of Research
Center on Rural
Economic Development

Dr. Adam D. Liska

Researcher
Center on Rural
Economic Development

**BROADBAND RESEARCH
EVOLVING!**

FEDERAL COMMUNICATIONS COMMISSION (FCC)



- Agency in charge of “everything” communications
- Their data landscape is evolving fast (now a **platform!**)

**CHALLENGE: HOW TO MANAGE
CONSTANT CHANGE UNDER**

FEDERAL COMMUNICATIONS COMMISSION (FCC)

	Form 477	National Broadband Map	BROADBAND DATA 2014 BUT
US Census Boundaries	2010	2020	• 2 volum dataset files
Type of recording	self declarative	self declarative	
Granularity	Census blocks	Locations	• No “roa barrier encodi values
Services	Mobile/Fixed	Mobile/Fixed	
Timeframe	2014-2021	2022 - Ongoing	
Releases	twice a	twice a	

1. One release can have multiple versions

BEHIND THE SCENE:

HOW IT STARTED: MANUEL EX



Data Download

[Data Spec. for BDC Public Data Downloads](#) [BDC Public Data API Specifications](#)

By State

By Provider

Challenges

Availability Data As Of

Dec 31, 2023 (latest)

Last updated 9/6/24

Summary

Type

Provider List

Provider Summary by Geography Type

Provider Summary - Fixed Broadband




Provider Summary - Mobile Broadband

State

Select...

Please select a state to view available downloads.

0:00 / 0:33

-  Repeat (56)
-  Repeat (??)
-  Error (clicks)

... AUTOMATION WITH CODE ...

```
1 # [...]
2 # Download all FCC data needed
3 all : dirs dl_file list_file
4
5 include config.mk
6
7 ## dirs:    Create specific dirs to
8             download data
9 dirs : config.mk
10      @pwd
11      @mkdir -p $(DATA_PATH)
12      echo "create $(DATA_PATH)"
13
14 ## dl_file: Create a list of files
15             available, download them
16 dl_file : $(CSV_SRC) get_csv_dl.R
17           Rscript get_csv_dl.R $(CSV_SRC)
18           Rscript list_file_dl.R
19           $(DATA_PATH) $(CSV_SRC)
20 .PHONY : dl_file
```

- New pro

**CHALLENGE
AND MAIN
CODE SO**

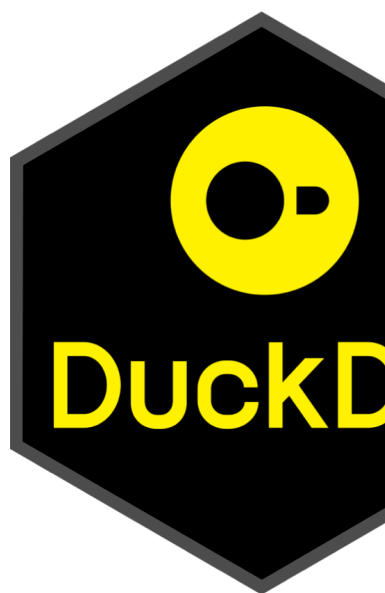
- Spoiler:
SQL

**CHALLENGE
ACCESSIBLE**

- Out of m
in DB

💡 **LIGHT FROM** POSIT::CONF

Workshop on DuckDB (Parquet / Arrow / OLAP)



SLACK IS BUZZING!

Warning

Contains real slack messages with typos!



Olivier Leroy 6:16 PM

I have the csv



I am stuck with my touch pad copy pasting in slack



DUCKDB SOLVED THE LACK OF REAC

... CODE PACKAGED

Visual

Code

```
library(cori.data.fcc)

dir <- "data_swamp/nbm/"

get_nbm_release()

system(sprintf("mkdir -p %s", dir))

dl_nbm(
  path_to_dl = "data_swamp/nbm",
  release_date = "June 30, 2023",
  data_type = "Fixed Broadband",
  data_category = "Nationwide",
)
```

- DuckD
- Reduce
- Comple
- place a
- so we c
- what b

CORI.DATA.FCC

WHAT IS PACKAGED CODE?

ONE WHITE LIE: CODE THAT WORKS ON
“MY” COMPUTER

BENEFITS FROM THE LANGUAGE ECOSYSTEM

- Tested against GitHub runner: it doesn't just work on my computer!
- [Website](#) to share it/distribute it

OPEN SOURCE: WHY SHARING



- Way to **publish**: we get more value
- Way to **share**: code is used!
- A place where we can **exchange** in a constructive way

HOW TO GET IT:

The source code is hosted on [GitHub](#) (Version

You need the R package `{remotes}`.

```
1 install.packages("remotes")
2 remotes::install_github("ruralinnovation/cori.data.fcc")
```

 Check the version you have and see what
offering 

```
1 packageVersion("cori.data.fcc")
```

```
[1] '0.0.1'
```

WHAT CAN IT DO: CHOOSE YOUR ADVENTURE!

- I need to go back to the source: Do you want data from FCC
- I need primary data but working with hundreds of CSV is not for me: Do you want data from CORI (NBM / Form 477)
- Census block is perfect for me: Do you want transformed data for NBM from Cook County)
- Guide you in those steps!

PACKAGING CODE = EXTRA TEAM MEMBER 🧑🏻‍🔧

we've talked about how packages can act like such as the **IT Guy**, **Analyst**, Tech Lead, or Product

[...] developed packages are [...] extra expert members.²

1. <https://www.emilyriederer.com/post/team-of-packages/#collaboration>
2. <https://milesmbain.xyz/posts/data-analysis-reuse/>

PACKAGING CODE = INCREASED INNOVATION / INSIGHT

- Functions designed to get data and insights in your environment
- **Capitalize on it!**
 - New project that is requiring **block estimates** -> add it!

EXAMPLES OF USES CAS

FIRST EXAMPLE: BLOCK COVER LIKE

Place holder for Camdem blog post

WHAT ARE THE ISP IN OHIO?

Visual

Code

Show

5 ▾

 entries

Search:

Provider Name	<div>Cnt</div> <div>Block</div>	<div>Cnt</div> <div>Rel.</div>
Cox Communications, Inc	2163	9414
Ottoville Mutual Telephone Company	270	992
Telephone Service Company	570	2516

Provider Name	Cnt Block	Cnt Rel.
SYCAMORE TELEPHONE COMPANY	404	1437
The Champaign Telephone Company	531	2287

Showing 1 to 5 of 121 entries

SUMMARY

- Primary data can give more power (in new field)!
- (Primary) data is only useful if you **effectively** use it.
- Use code that are abstracting the
- Easier to **capitalize/build** on it: get less tech debt
- we need those extra “team members”

CONTACTS

Website: <https://ruralinnovation.us>

[LinkedIn](#) | [Twitter](#) | [Facebook](#) | [Instagram](#) | [YouTube](#)