Curtin University

THE CURTIN INSTITUTE FOR DATA SCIENCE

Computational Skills Training

HDR Induction for Supervisors and Students

July 2024

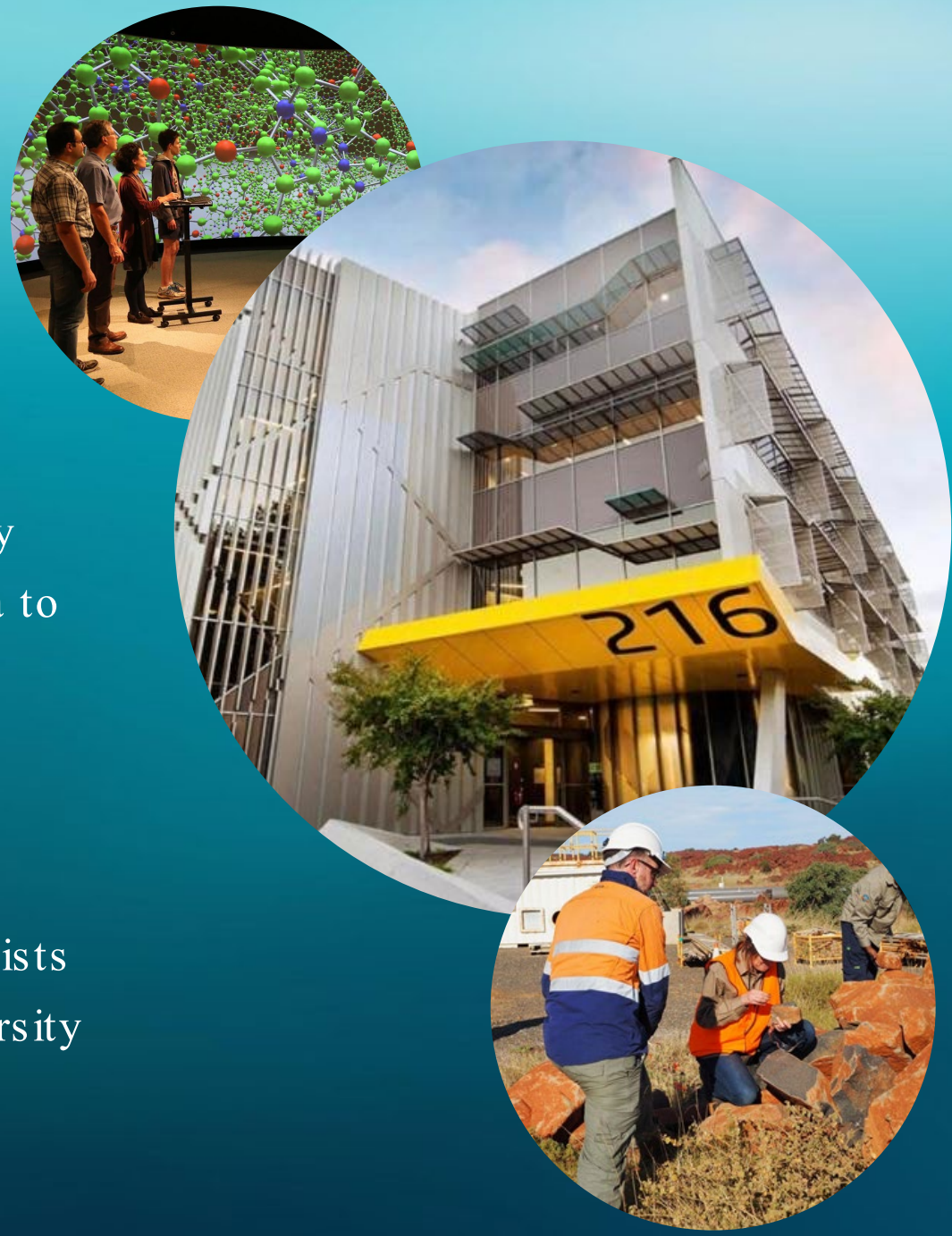
Dr Dan Marrable
Dr Kathryn Napier

Curtin Institute for Data Science Overview

Australia's largest university-based data science research institute.

The Curtin Institute for Data Science is an interdisciplinary knowledge accelerator providing innovative solutions in data to solve real-world challenges.

- Established in 2016
- Based at Curtin University, Bentley Campus, Western Australia
- Multi-disciplinary team of 40+ in-house technical specialists
- Network of 200+ virtual researchers across Curtin University
- Data projects across all domains



Curtin University

THE CURTIN INSTITUTE FOR DATA SCIENCE

PROFESSOR MELANIE JOHNSTON-HOLLITT

Director, Curtin Institute for Data Science

Director, Australian Space Data Analysis Facility

Curtin University

THE CURTIN INSTITUTE FOR DATA SCIENCE

Our Specialities

The Institute provides end-to-end data support across nine specialties:

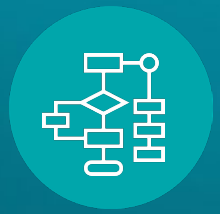
Big Data Analytics

Machine Learning and Artificial Intelligence

High Performance Computing

Research Software Development

Data Engineering



Modelling and Optimisation

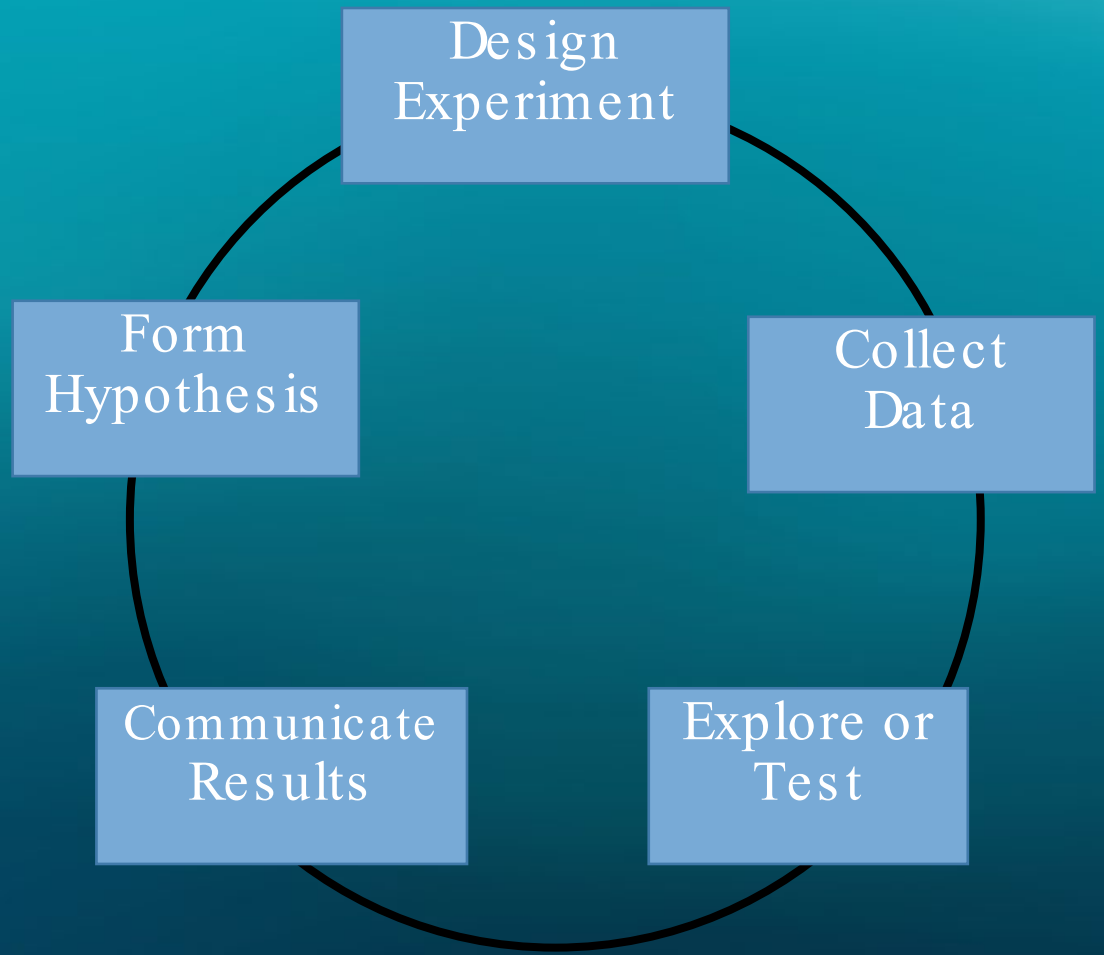
Visulisation and Simulation

Education and Training

Project Scoping and Life Cycle Management

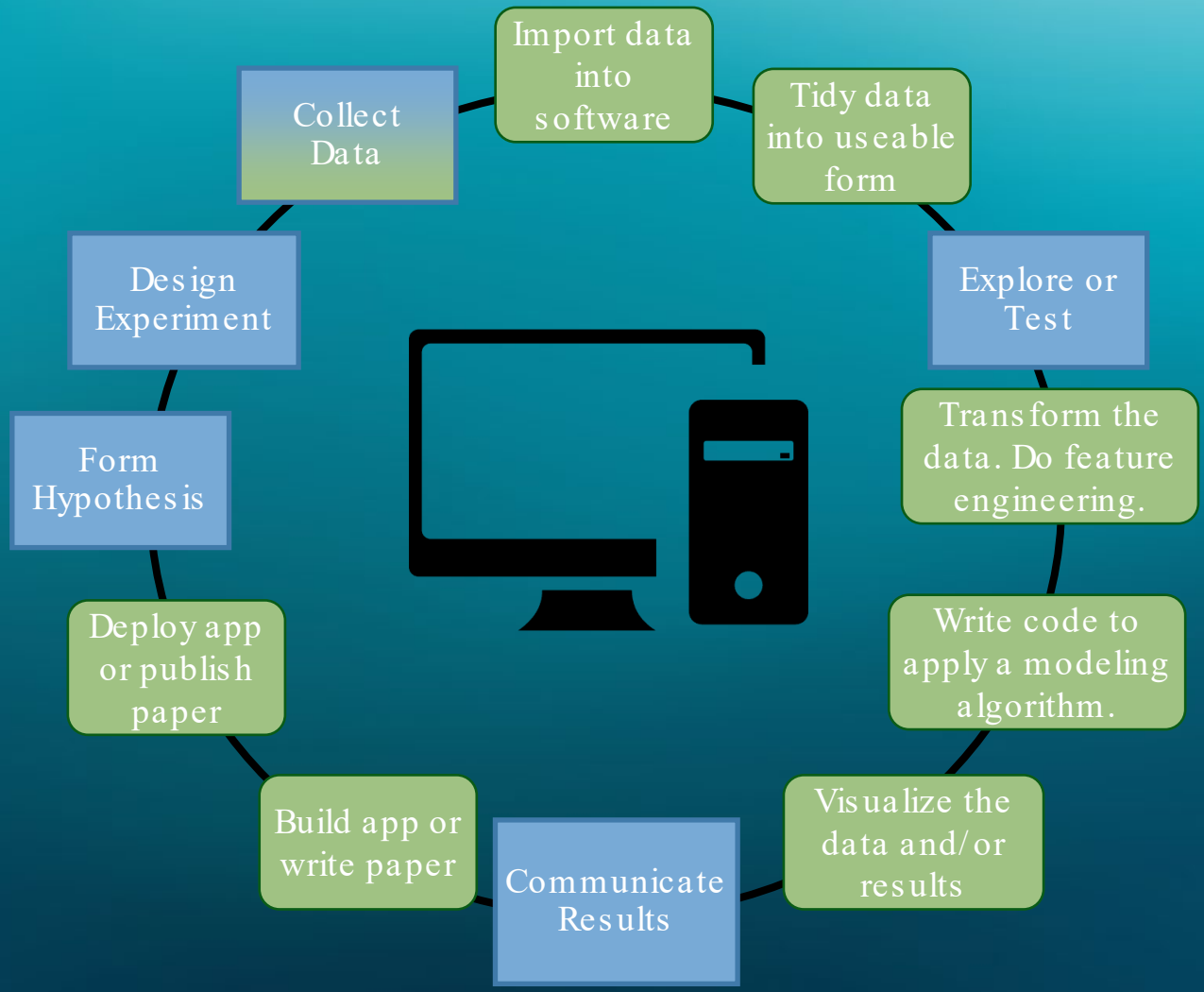
The importance of software in research and academia

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

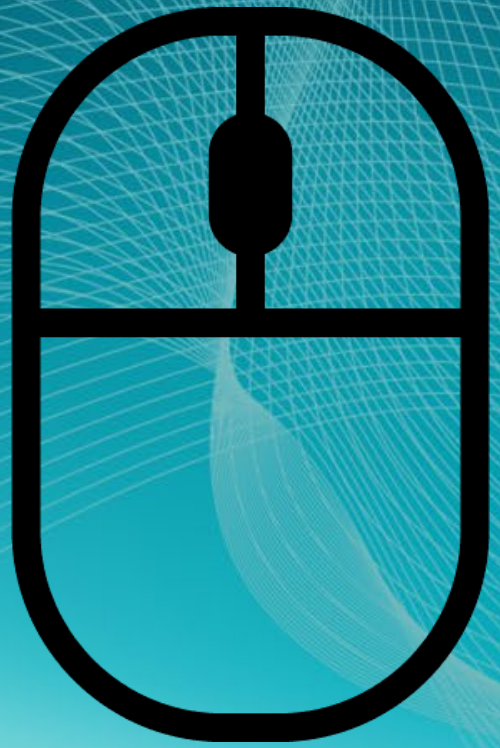


The importance of software in research and academia

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE



What does your data analysis workflow look like?



Point and click?

Ok for simple tasks

One at a time…

Hard to automate or reproduce

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

A computer language for scientists

Human thought → Machine language

CC BY-SA RStudio

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

A computer language for scientists

High level/ interpreted languages, Instrument Software GUI, Search Engines, API, etc

C, C++, FORTRAN, Instrument Software, etc

Human thought

Machine language

CC BY-SA RStudio

Why use R and / or Python?

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

Python and R are supported by a large open source community

Why use R and / or Python?

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

Questions tagged [python]

Ask Question

Python is a multi-paradigm, dynamically typed, multi-purpose programming language. It is designed to be quick to learn, understand, and use, and enforces a clean and uniform syntax. Please note that Python 2 is officially out of support as of 2020-01-01. For version-specific Python questions, add the [python-2.7] or [python-3.x] tag. When using a Python variant (e.g. Jython, PyPy) or library (e.g. Pandas, NumPy), please include it in the tags.

Learn more... Top users Synonyms (4)

2,101,829 questions

Newest | Active | Bountied 25 | Unanswered | More | Filter

0 votes
0 answers
4 views
Add note with attachments on Freshservice
I am trying to add a note with an attachment to a Fresh
working. Below is my code: # * Construct the Freshserv
python python-requests multipartform-data fresl

0 votes
0 answers
4 views
saisonal period m in SARIMA for 15min-data
I have data of the strength of the sun's rays every 15 m
make some predictions with SARIMA / SARIMAX. Now I
python arima sarimax

Questions tagged [r]

Ask Question

R is a free, open-source programming language & software environment for statistical computing, bioinformatics, visualization & general computing. Please use minimal reproducible examples others can run using copy & paste. Show desired output entirely. Use dput() for data & specify all non-base packages with library(). Don't embed pictures for data or code, use indented code blocks instead. For statistics questions, use https://stats.stackexchange.com.

Learn more... Top users Synonyms (3)

480,105 questions

Newest | Active | Bountied 12

0 votes
1 answer
10 views
Can I joint 2 dataframes in R by an ID column and
I am trying to join common ID values in 2 dataframes to their
have a list with demographic data and another with blood san
r dplyr lubridate

0 votes
0 answers
7 views
Ranking changes according to their frequency (R
enter image description here I am working on a school projec
based on data from an Excel table and this is one of them D is
r rstudio

Questions tagged [spss]

Ask Question

SPSS is a statistics package. Originally released in 1968, it is currently owned by IBM.

Learn more... Top users Synonyms

2,031 questions

Newest | Active | Bountied | Unanswered | More | Filter

0 votes
0 answers
16 views
3-way ANOVA pairwise comparisons or t-tests? (SPSS)
I am investigating differences between written evaluations about men and women. For each sentence in each evaluation the valence (positive/negative) and performance-relatedness (yes/no)....
spss anova t-test pairwise
bobi 1 asked Feb 1 at 12:04

Why use R and / or Python?

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

Python and R are free and open source

No paywalls
No subscription costs
You can use your favourite OS
Open source software is free (as in freedom)

...everyone can afford their vegetables this week!

Why use R and / or Python?

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

The code is readable

It is your record of your analysis!

...research becomes reproducible

JAVA

```java
public class Main {
  public static void main(String[] args) {
    System.out.println("hello world");
  }
}
```

PYTHON

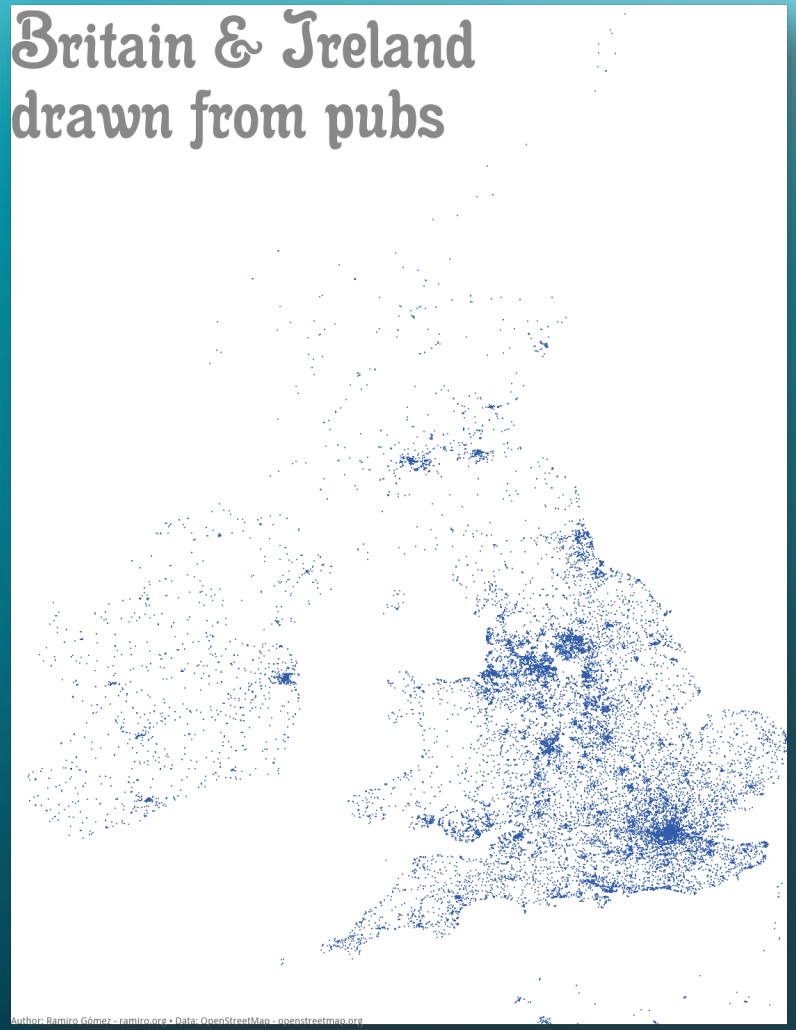
```python
print('hello world')
```

Why use R and / or Python?

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

Analysis is scalable

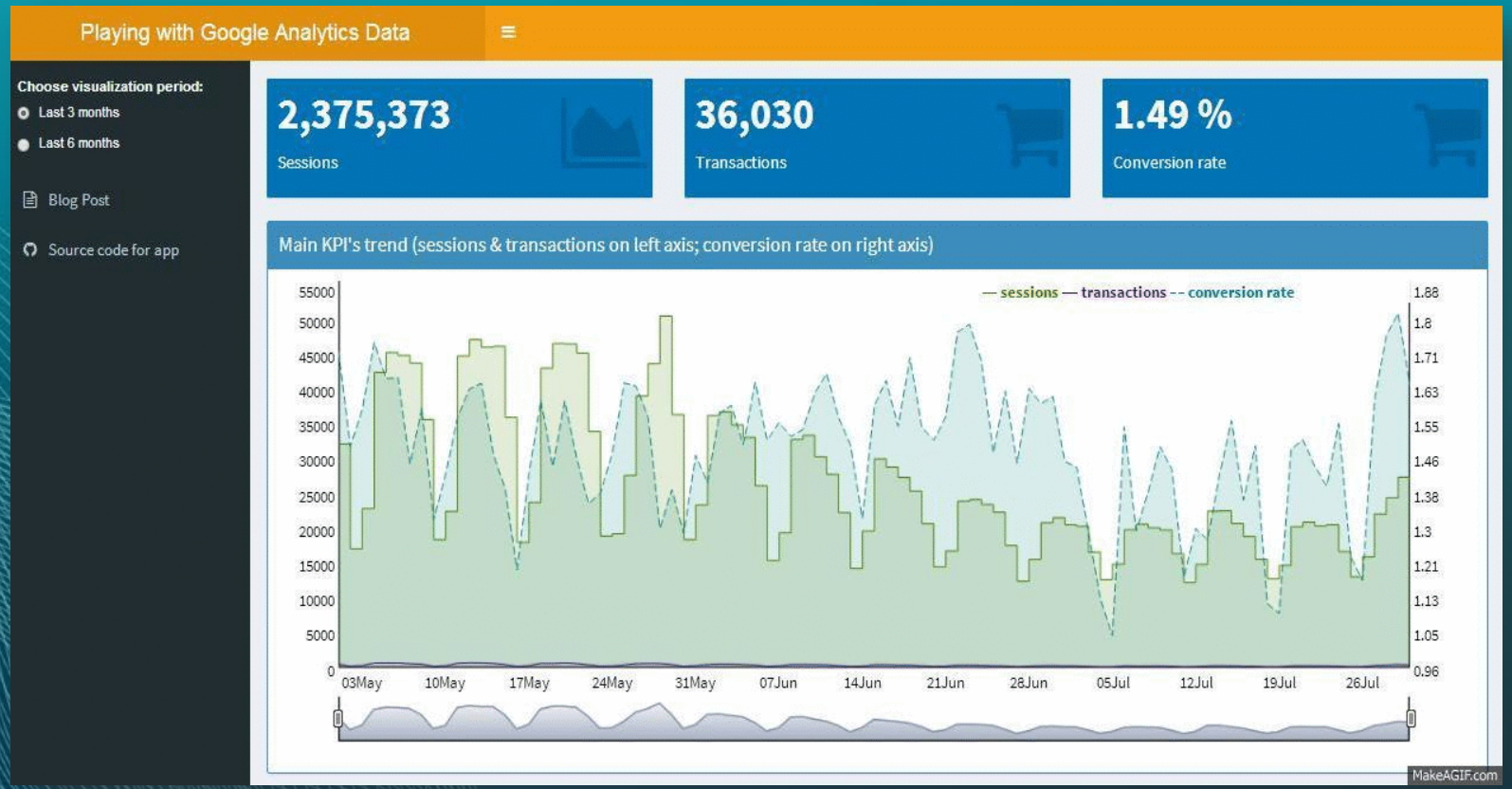
```
fig = plt.figure(figsize=(20, 30))
markersize = 1
markertype = ','  # pixel
markercolor = '#325CA9'  # blue
markeralpha = .8 #  a bit of transparency

m = Basemap(
    projection='mill', lon_0=bbox['lon'], lat_0=bbox['lat'],
    llcrnrlon=bbox['ll_lon'], llcrnrlat=bbox['ll_lat'],
    urcrnrlon=bbox['ur_lon'], urcrnrlat=bbox['ur_lat'])

# Avoid border around map.
m.drawmapboundary(fill_color='#ffffff', linewidth=.0)

# Convert locations to x/y coordinates and plot them as dots.
lons, lats = zip(*locations)
x, y = m(lons, lats)
m.scatter(x, y, markersize, marker=markertype, color=markercolor, alpha=markeralpha)

# Set the map title.
plt.annotate('Britain & Ireland\ndrawn from pubs',
             xy=(0, .87),
             size=120,
             xycoords='axes fraction',
             color='#888888',
             family='Gloria')

# Set the map footer.
plt.annotate('Author: Ramiro Gómez - ramiro.org • Data: OpenStreetMap - openstreetmap.org',
             xy=(0, 0),
             size=14,
             xycoords='axes fraction',
             color='#666666',
             family='Droid Sans')

plt.savefig('img/britain-ireland-drawn-from-pubs.png', bbox_inches='tight')
```



http://ramiro.org/notebook/mapping-pubs/

Why use R and / or Python?

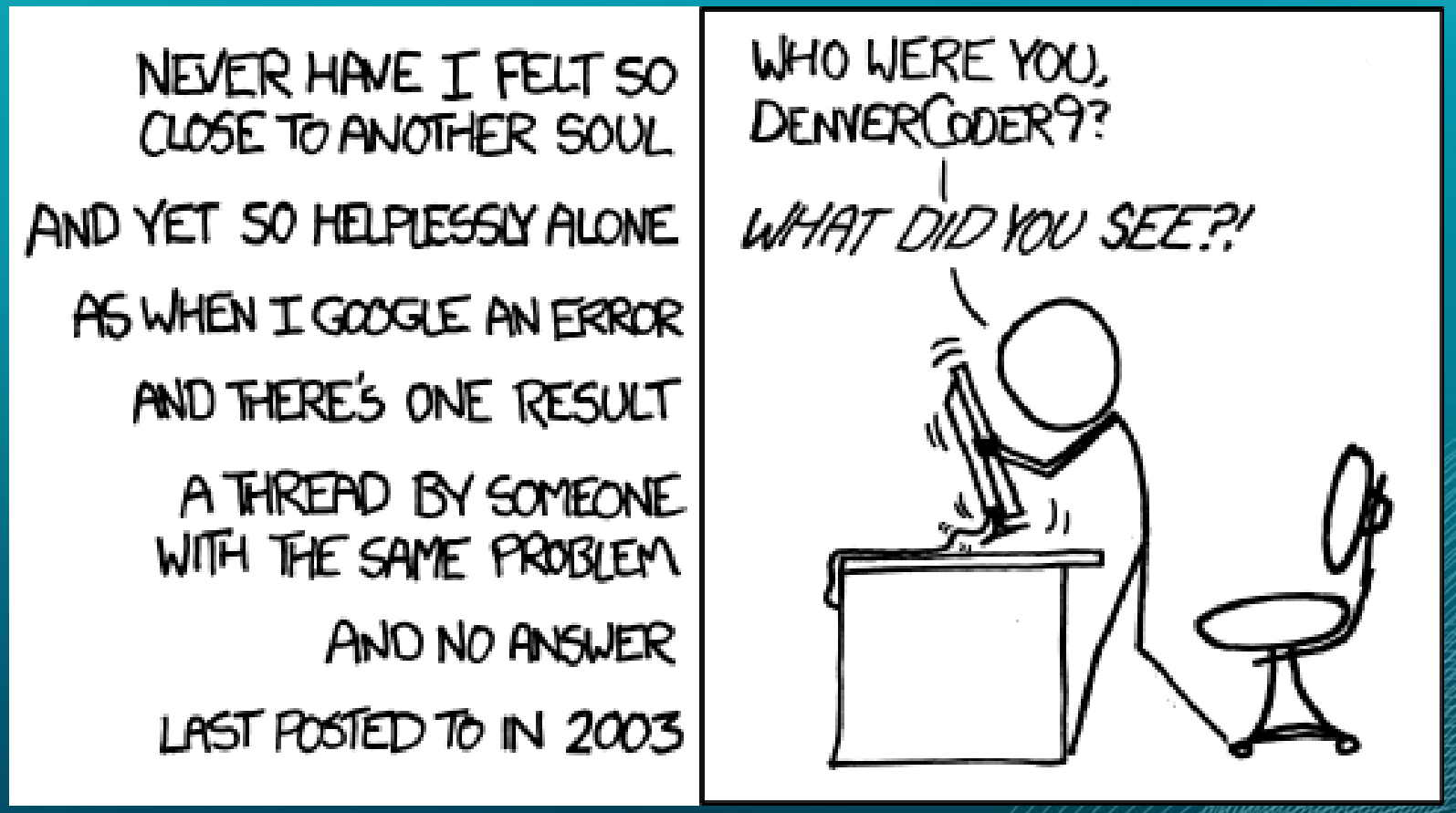
Results are shareable



https://www.r-bloggers.com/playing-with-r-shiny-dashboard-and-google-analytics-data/

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

Learning to code on your own != trivial!

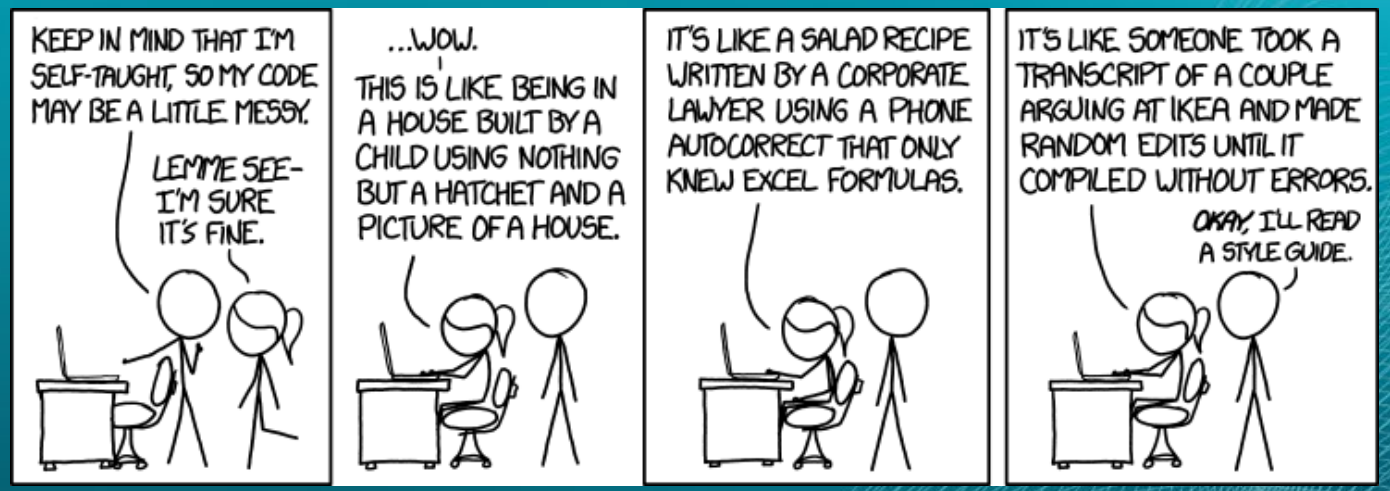


Leveraging the community for skills training

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

Teaching

Carpentries training



Ongoing support

CIDS Book a Technical Expert Sessions

Code review sessions in your area

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

The Carpentries

THE CARPENTRIES

Teach foundational computational, coding, and data science skills to researchers

Workshops are in-person events, generally two days long

Global community of volunteers

Carpentry certified instructors are trained in pedagogy

Lessons are freely available under the Creative Commons - Attribution License

Curtin University
THE CURTIN INSTITUTE FOR DATA SCIENCE

Support from the Institute

CIDS Book a Technical Expert Sessions

Held every Tuesday, 3pm~~-~~4pm at the CIDS office (B216:206/207)- email **curtinids@curtin.edu.au** to rsvp, **click here for more information.**

Computing Workshops

The Carpentries

- Python, R and Git workshops twice a year - **Sept 4-6 2024 (python and Git)**
- **Join our mailing list here** to be informed when registrations open- they fill out fast!

@CurtinIC

Follow us on linkedin

datascience.curtin.edu.au
Join our mailing list and Slack!

curtinids@curtin.edu.au