Open Learning Education

Exploratory Data Analysis

Contributors : Michelle Manning, Weigeng Li & Zutima Tuladhar
DACSS 604 : Final Project Submission
University of Massachusetts Amherst

Contents

A. Introduction OLE

B. The Audiences

C. The Questions

D. The Data

E. Results

F. Summary of Findings

G. Recommendations and Reproducibility

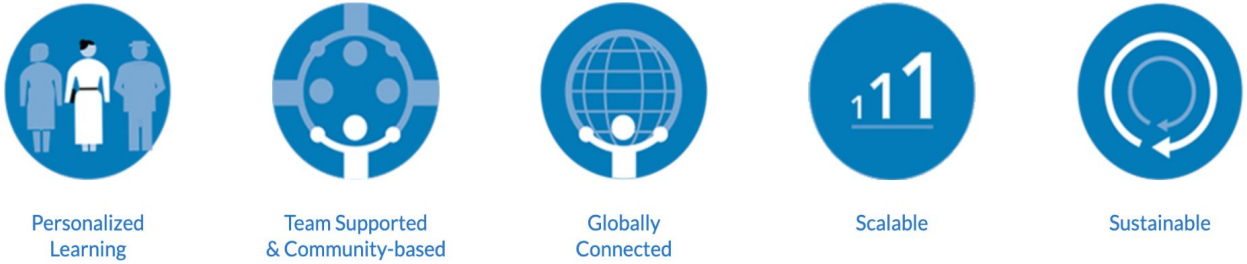
Open Learning Exchange

Moving from education to learning

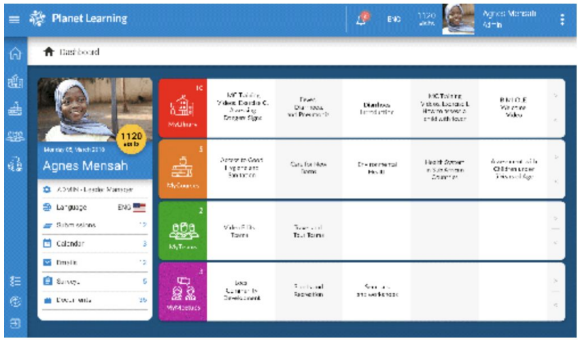
Let's talk about OLE :

Name : Open Learning Education

Locations: Peru, Mexico, United States, Canada, Ghana, Togo, DRC, Uganda, Kenya, Cameroon, Equatorial Guinea, Somalia, Madagascar, Turkey, Bulgaria, Lebanon, Jordan, Nepal, India, and Cambodia.



What are their methods?



Tool :

- The Raspberry Pi, costing US$35. is often used as the server.
- It can easily be powered locally by batteries and solar cells.
- The complete system, tablets, server, battery, camera can be contained in a small wheeled suitcase, or backpack, moved from one location to another, up and running in less than a minute.

User Indicators

- Activity reports are aggregated by gender and age and detail members use
- the number of resources opened
- the names of the most frequently opened resource, and member resources ratings.

Process & Timeline

First Steps and Preliminary Findings

- **First Meeting with Dr. Meredith Rolfe and Maddie Hertz**

- **Initial Data Codes and Cleaning (Initial Github)**

- **First Meeting with Dr. Richard Rowe (CEO and Chairman)**

First Steps and Preliminary Findings

- **First Meeting with Dr. Meredith Rolfe and Maddie Hertz**

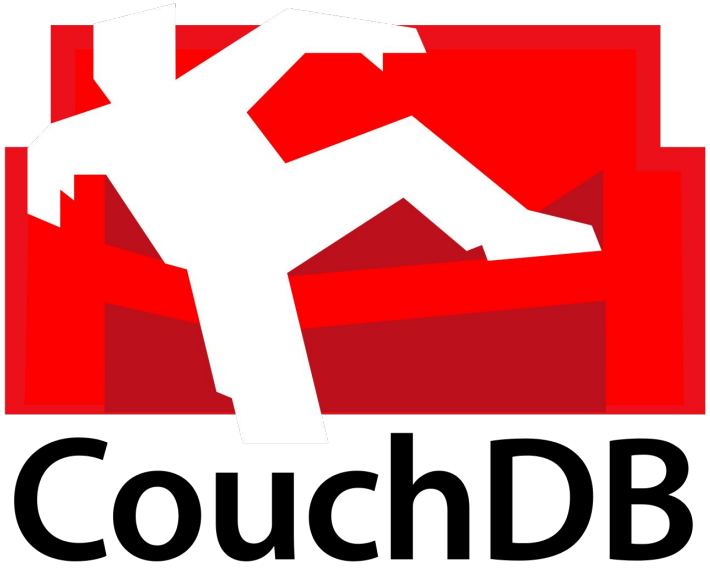
- **Initial Data Codes and Cleaning (Initial Github)**

- **First Meeting with Dr. Richard Rowe (CEO and Chairman)**

Technical Background

OLE use CouchDB

- NoSQL Database
- Non-relational database
- Each database is a collection of independent documents (key- value pairs)
- Documents do not have to contains same keys



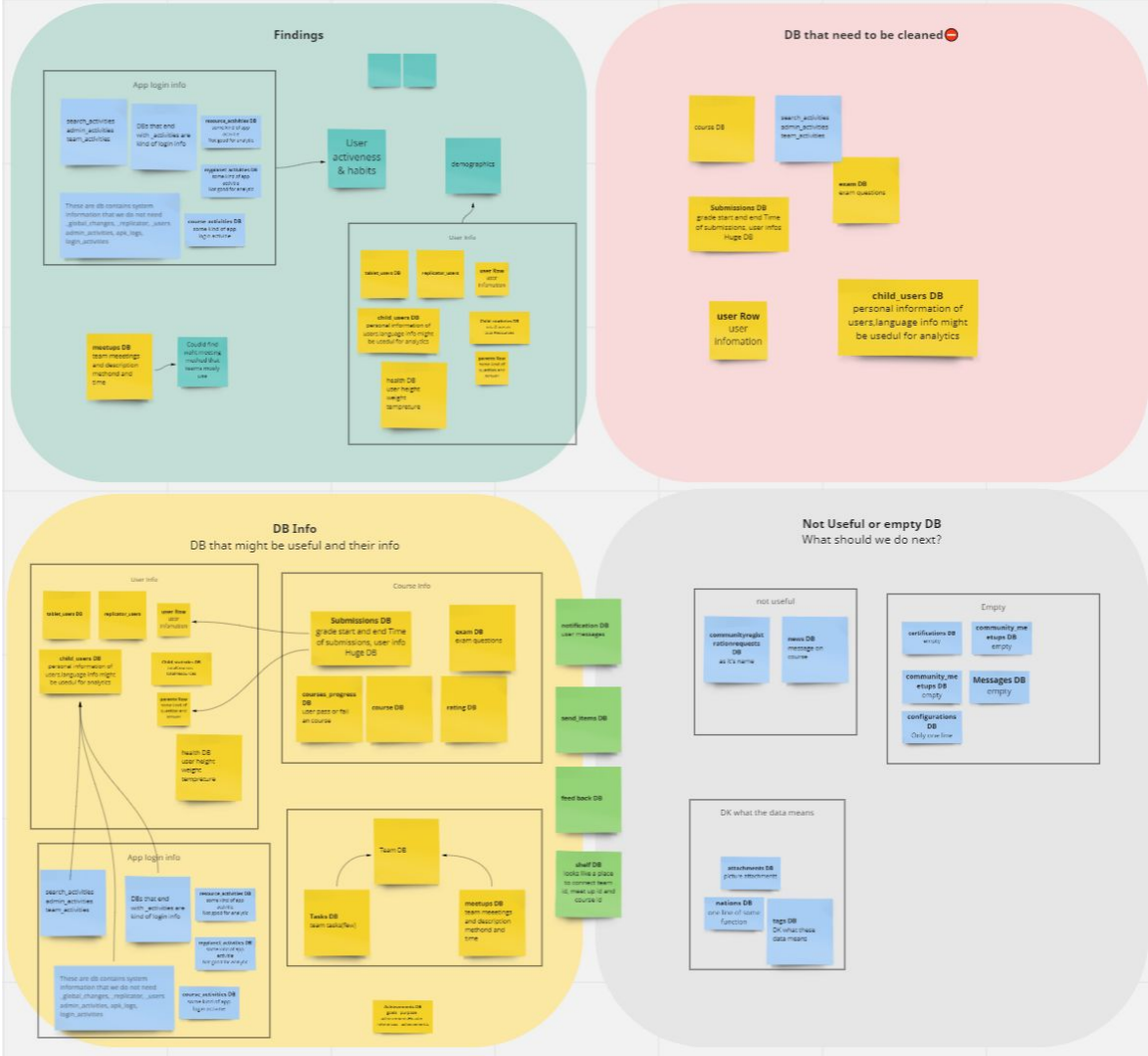
Initial Data Codes – Start with a soup

```r
#flatten into a dataframe of documents
docids<-map_dfr(docs$rows, flatten_dfc)
soup <-map(docids$id[1:10], ~doc_get(ole, database, .x))
```

Name	Type	Value
soup	list [10]	List of length 10
[[1]]	list [8]	List of length 8
[[2]]	list [13]	List of length 13
[[3]]	list [9]	List of length 9
[[4]]	list [13]	List of length 13
[[5]]	list [13]	List of length 13
[[6]]	list [11]	List of length 11

Database Exploring

- Build Relationship Diagram
- Findings and possible research subject
- Assign data cleaning Tasks



Data cleaning

- Encoding
- Transform Timestamp to somalia local time
- Calculate age, generation and user time spent
- Remove outliers and wrong data
- Merge Databases

Cute Childrens

level	type	gender	phoneNumber
	user		
Beginner	user	female	23¾

level	type	gender	phoneNumber
Beginner	user	female	355
Beginner	user	female	36999
Beginner	user	female	23555
Beginner	user	female	56677⅞
Intermediate	user	female	000
Beginner	user	female	6555
Advanced	user	female	[redacted]@ole.org

Wrong Data and Outliers

Some people really love this app (using this app for more than a month)

```{r}
login_activities[which(usage_sec>3600000),]
```

Description: df [88 × 9]

parentCode <chr>	loginTime_day <chr>	loginTime_time <chr>	logoutTime_day <chr>	logoutTime_time <chr>	usage_sec <dbl>
somalia	05-27-2019	11:45:04	07-09-2019	16:33:01	3732477
somalia	05-27-2019	12:20:54	07-09-2019	16:33:41	3730367
somalia	11-14-2020	15:51:08	12-26-2020	20:50:48	3646780
somalia	06-11-2019	15:41:39	07-30-2019	14:14:06	4228347

Get that Time traveler! (people logout before login)

```{r}
login_activities[which(usage_sec<0),]
```

Description: df [45 × 9]

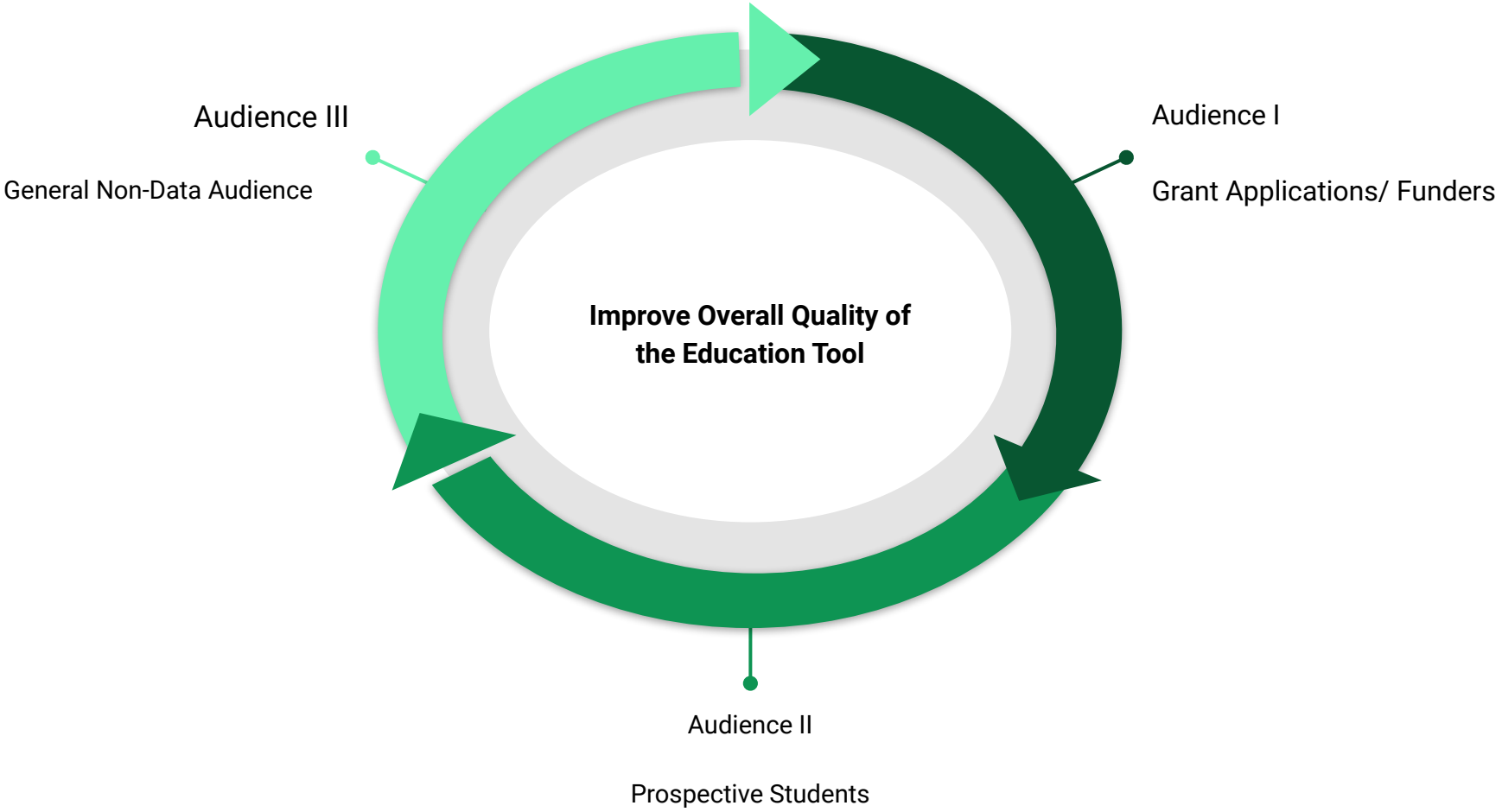
parentCode <chr>	loginTime_day <chr>	loginTime_time <chr>	logoutTime_day <chr>	logoutTime_time <chr>	usage_sec <dbl>
somalia	05-04-2020	22:26:27	05-04-2020	13:01:59	-33868
somalia	04-23-2019	20:27:43	04-17-2019	17:35:34	-528729
somalia	08-30-2018	13:07:03	04-30-2017	4:50:54	-42106569
somalia	05-04-2019	9:17:19	05-04-2019	9:08:39	-520
somalia	05-04-2019	9:43:59	05-04-2019	9:07:45	-2174
somalia	03-07-2020	14:30:31	03-04-2020	13:37:17	-262394
somalia	03-07-2020	15:27:44	03-04-2020	13:50:20	-265044
somalia	09-11-2018	10:54:18	05-01-2017	5:13:08	-43047670
somalia	09-03-2018	16:50:27	01-05-2010	8:35:30	-273309297
somalia	11-05-2018	15:29:58	11-05-2018	15:26:50	-188

Data Summary

- Users: Over 2,300 (56% Children and Youth, 45% adult)
- Number of courses: 69
- App Login: Over 30,000 record
- Accessing recourse: About 50,000 Times
- Class Attendance: About 13,000 Times
- Server Usage: Over 26,000 record

First Steps and Preliminary Findings

- **First Meeting with Dr. Meredith Rolfe and Maddie Hertz**
- **Initial Data Codes and Cleaning (Initial Github)**
- **First Meeting with Dr. Richard Rowe (CEO and Chairman)**


Audience III
General Non-Data Audience
Audience I
Grant Applications/ Funders
Improve Overall Quality of the Education Tool
Audience II
Prospective Students

Questions We Seek To Answer

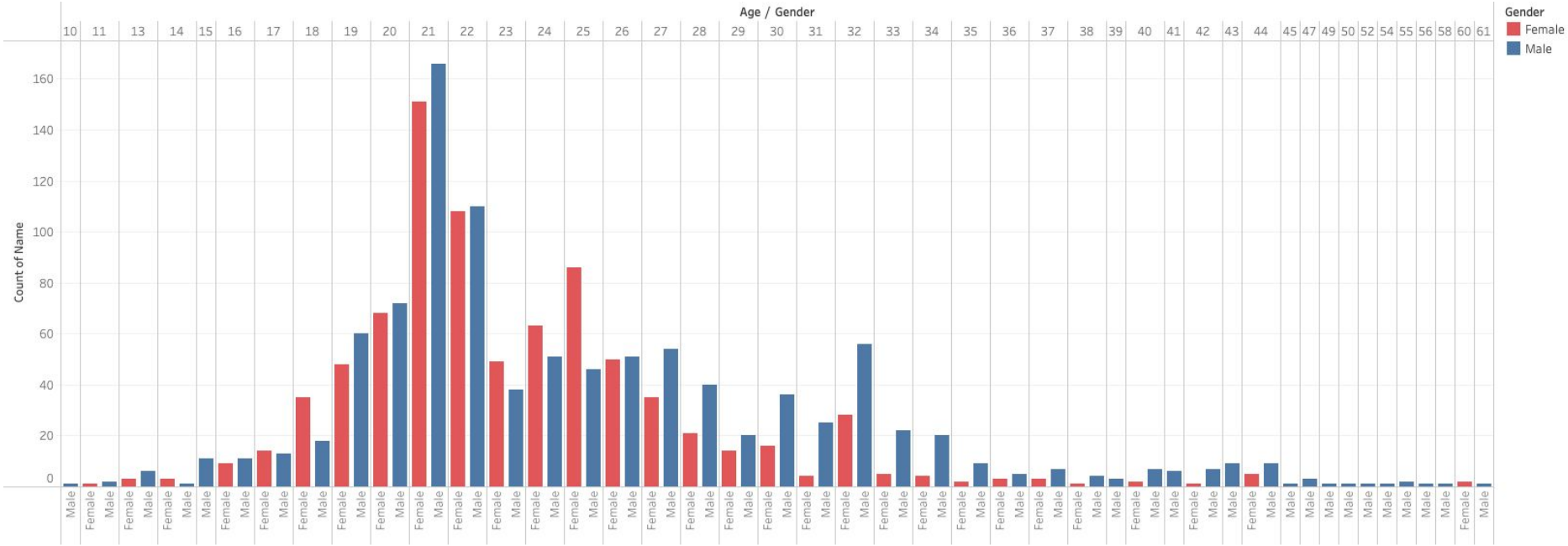
A. *How does gender influence app/course use?*

B. *What is the age demographic of the learners?*

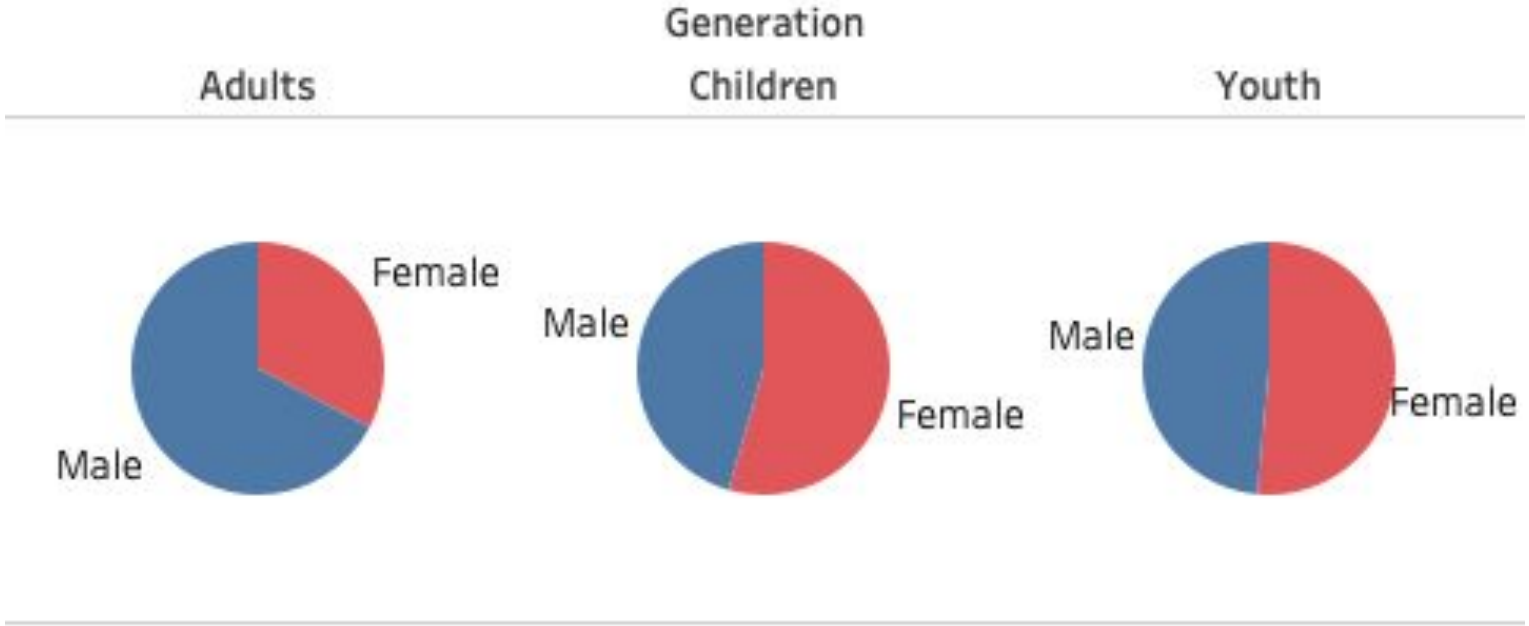
C. *Are there any patterns in course activity?*

Data Visualization

Age & Gender

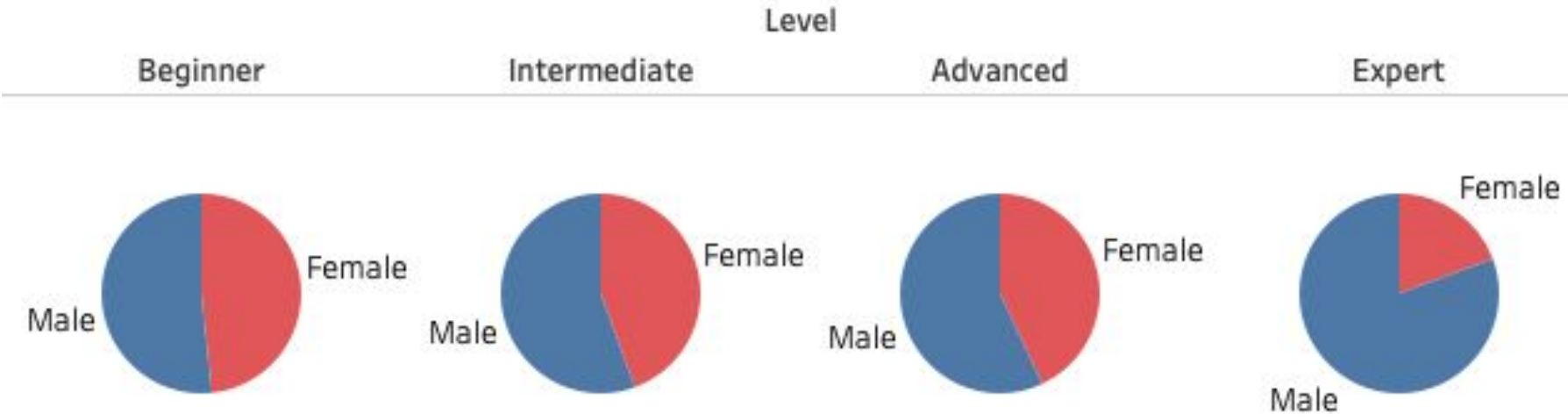

Age / Gender
10 11 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 47 49 50 52 54 55 56 58 60 61
Count of Name
0 20 40 60 80 100 120 140 160
Male Female Male Female Male Female

Generation by Gender



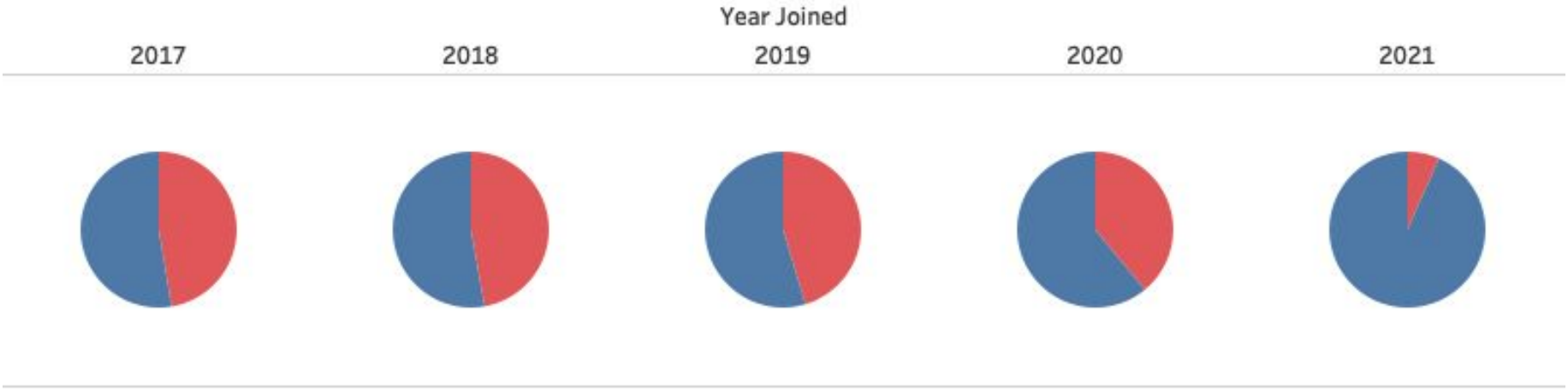
Gender (color) broken down by Generation. The view is filtered on Generation and Gender. Generations are defined as Adults: <24, Youth: 15-24, Children: <15. The Generation filter excludes NA and Null. The Gender filter keeps Female and Male.

Student Level by Gender



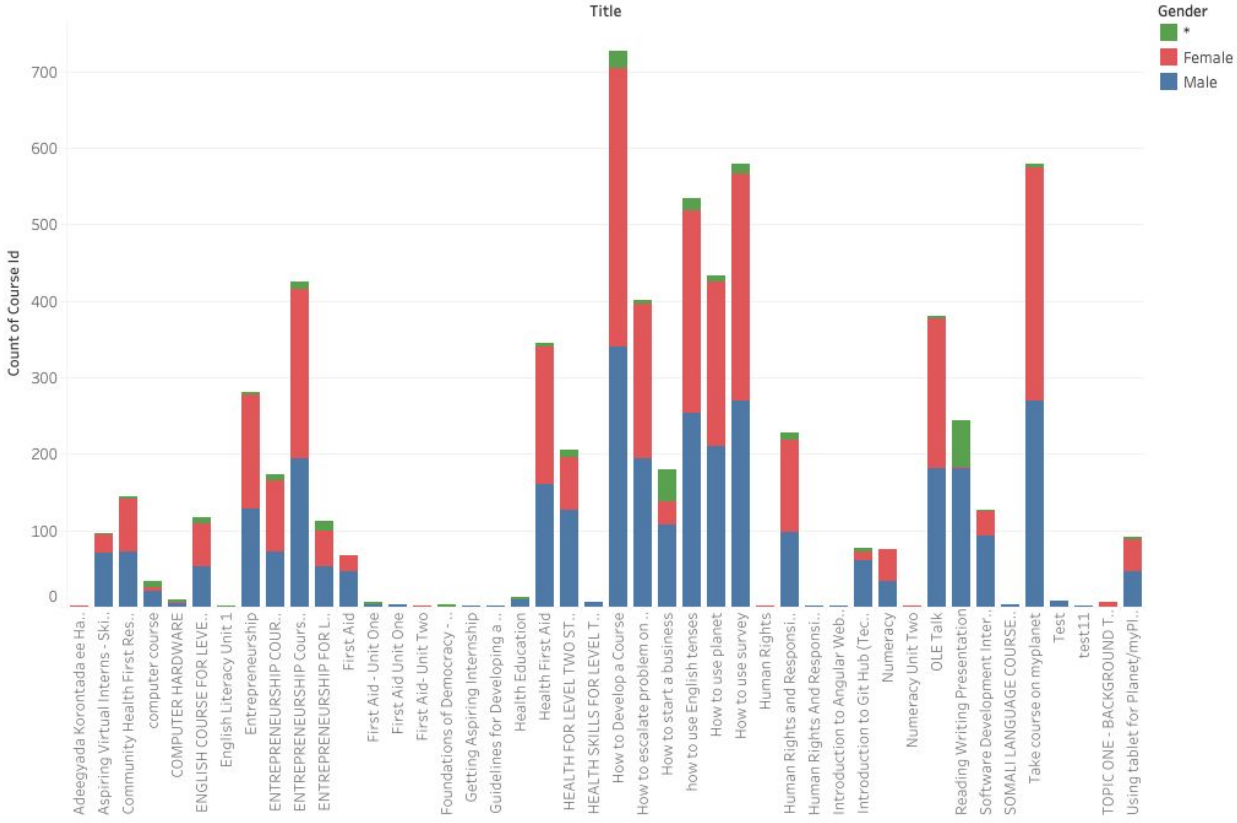
Gender broken down by Level. Color shows details about Gender. The marks are labeled by Gender. The view is filtered on Level, which excludes 1, NA and Null.

Year Joined by Gender



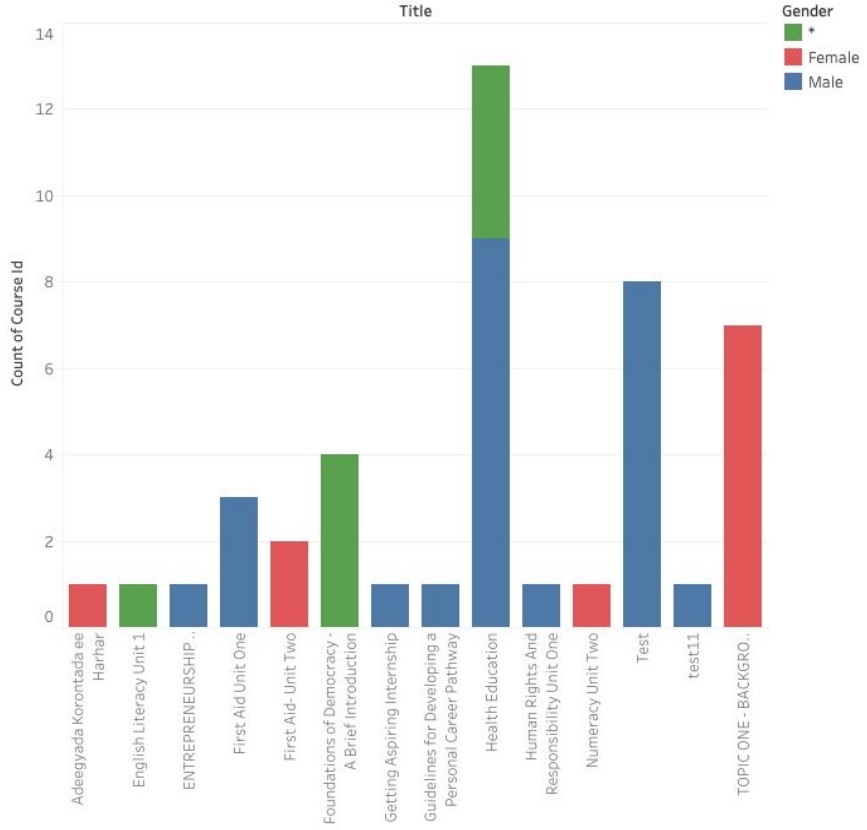
Gender (color) broken down by Year Joined Year. The view is filtered on Gender and Year Joined Year. The Gender filter keeps Female and Male. The Year Joined Year filter has multiple members selected.

Course Names by Gender



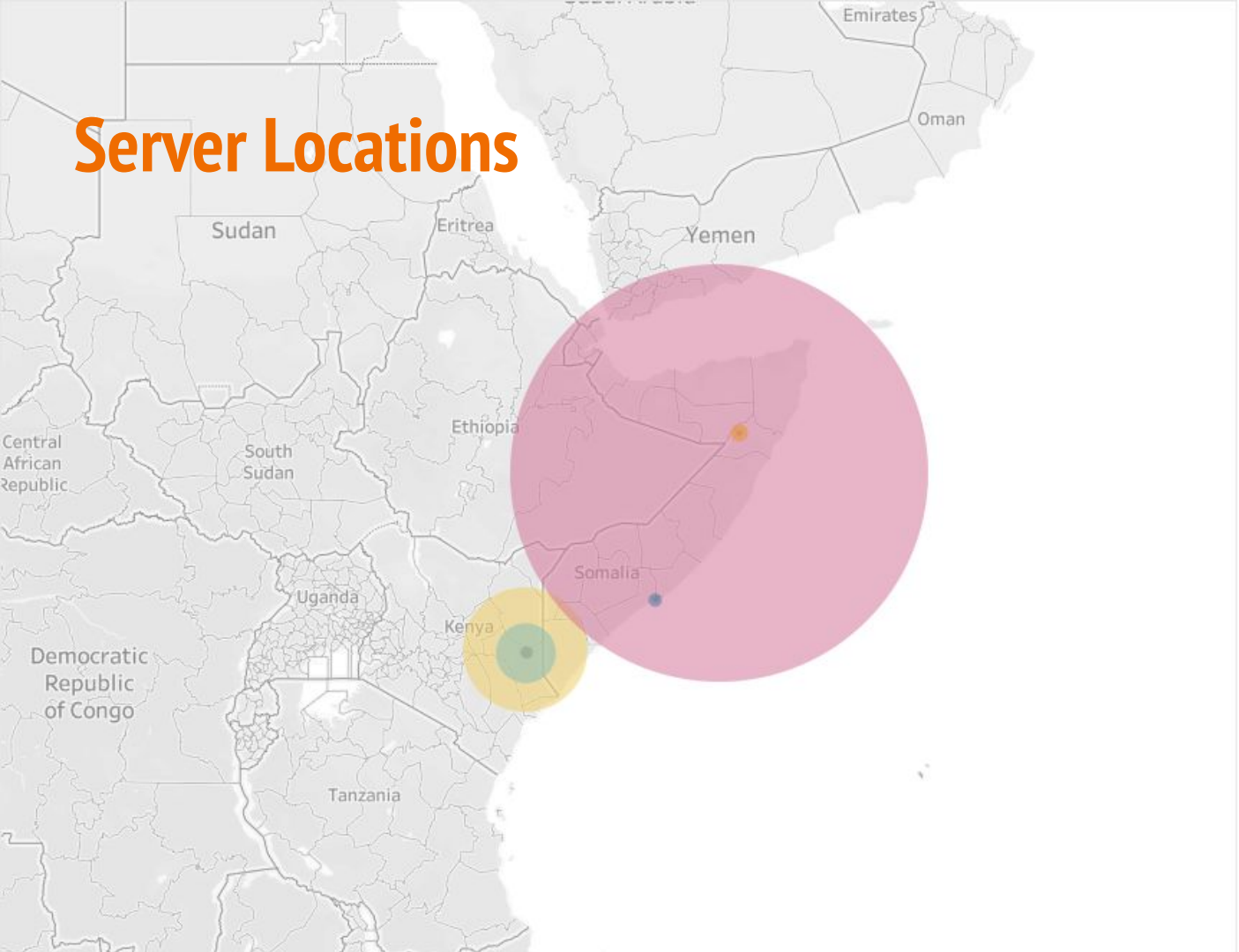
Count of Course Id for each Title. Color shows details about Gender (all_user.csv (Multiple Connections)). The view is filtered on Gender (all_user.csv (Multiple Connections)), which keeps Female and Male.

Course by Gender Outliers



Count of Course Id for each Title. Color shows details about Gender (all_user.csv (Multiple Connections)). The data is filtered on Course Id (course_activities_user), which keeps 19 of 146 members. The view is filtered on Gender (all_user.csv (Multiple Connections)), which keeps Female and Male.

Server Locations



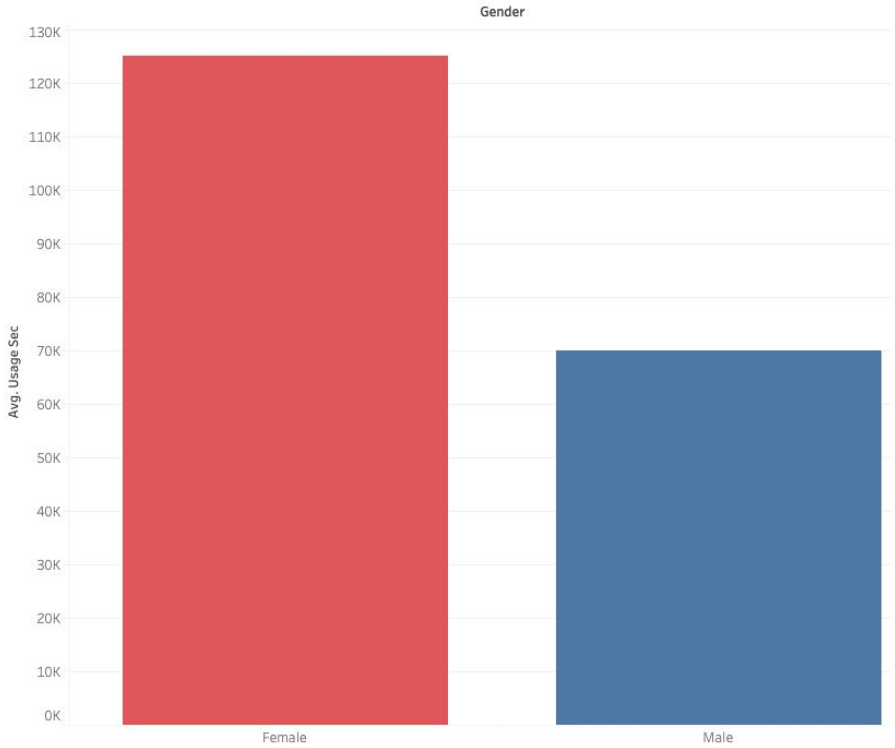
Count of Created On

5
5,000
10,000
15,000
22,211

Created On

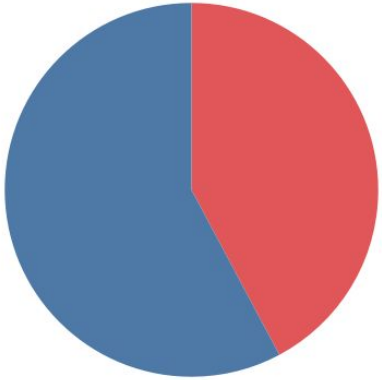
- abdirahman
- peacebuilder2
- peacebuilder5
- peacebuilder6
- peacebuilder7
- peacebuilder8
- peacebuilder9
- peacebuilder13
- peacebuilder14
- peacebuilder15

Time Spent on Courses



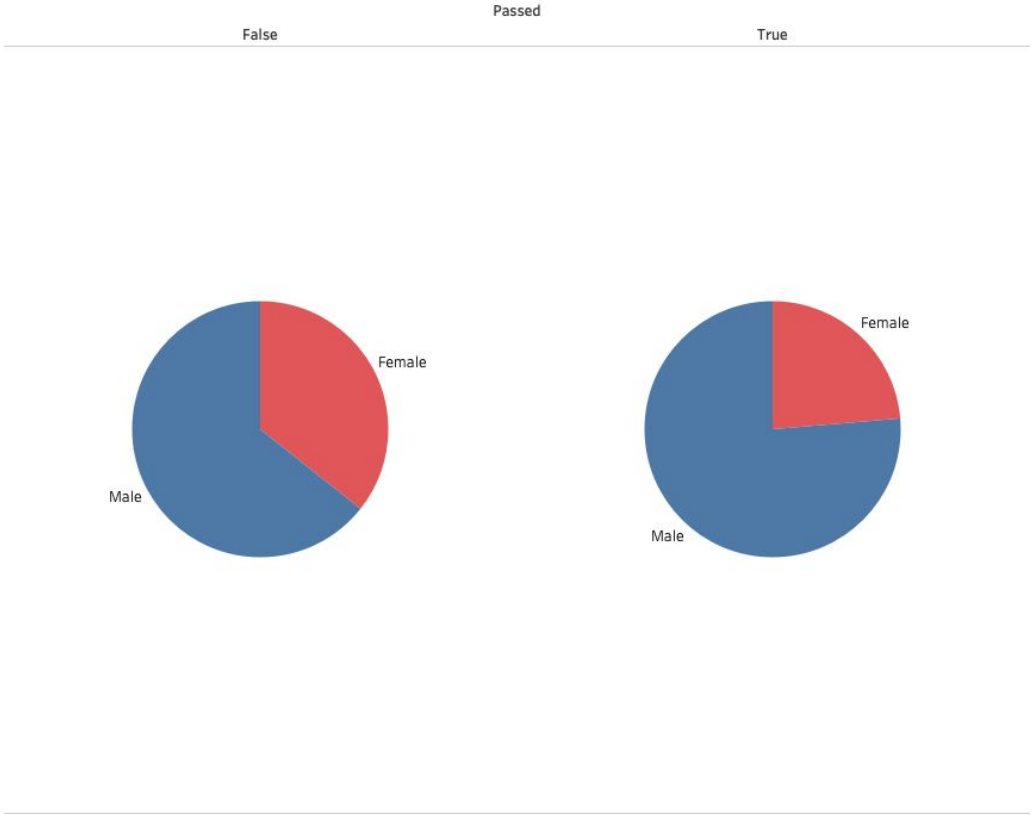
Average of Usage Sec for each Gender. Color shows details about Gender. The view is filtered on Gender, which keeps Female and Male.

Visit Count



Gender (color). The view is filtered on Gender, which keeps Female and Male.

Course Completion



Gender broken down by Passed. Color shows details about Gender. The marks are labeled by Gender. The view is filtered on Gender, which keeps Female and Male.

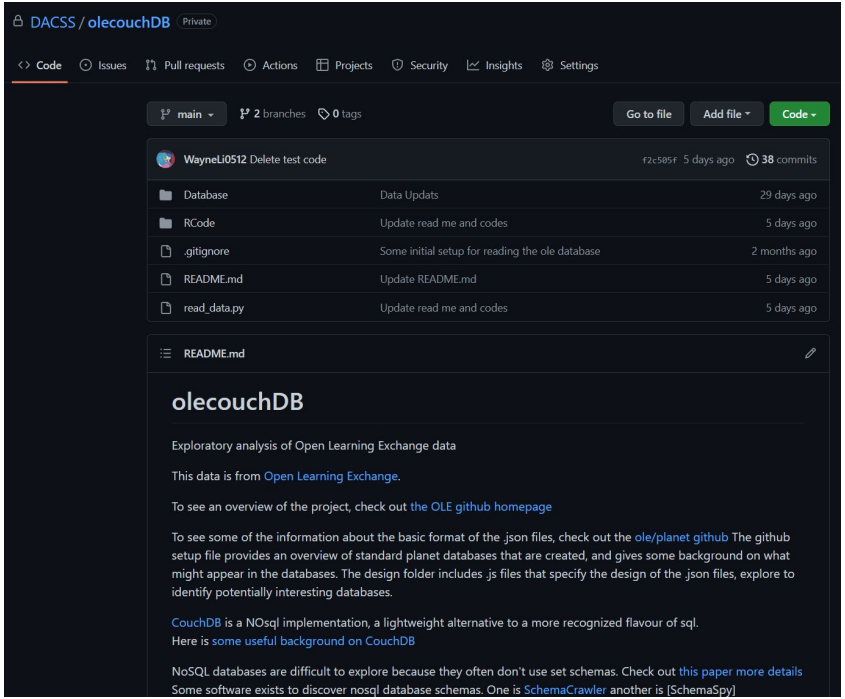
Final Findings & Suggestions

Summary of Findings

- People who are in their 20s are you main age demographic
- Overall, there are more men than women users
- Could support women learners as they are the minority at every student level and become a smaller proportion as level increased
- This could be done by focusing on the courses women are applying to and supporting women's issues
- Men tend to visit more often, but women on average spend more time on the courses
- Most of the users are using the peacebuilder 5 server, which is near Galkayo, Somalia.

Reproducibility and Collaboration

- Readme
- Code Comment (Google Style)
- Google Docs
- Github Repo



Thank You!