

FBN Data Report #1

8/23/2023

Salome “Sam” Grasland
samgrasland@gmail.com





Storing and Managing Data



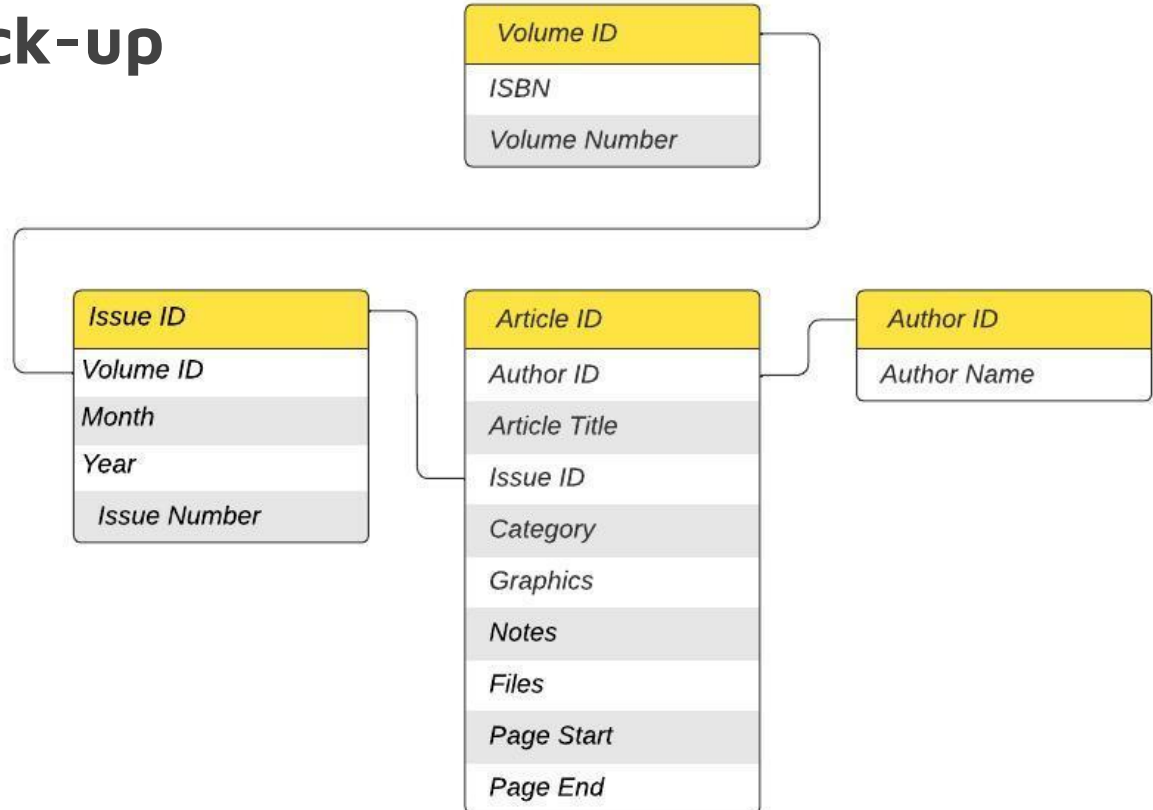
Data Entry and Cleaning

Currently FBN data is being stored and managed on a google sheet. The pros are that this data is easily accessible to a wide user base.. However, eventually this google sheet will result in unorganized data. For example, to the right I have pasted an image currently showing all the iterations of month in the database. Scanning over the list we can see several issues. The data for month includes input errors like spelling errors, abbreviations, or inconsistent formatting. While this can be cleaned it does add an extra element of processing time.

Month
Null
3
?
(Sep/Oct?)
April
August
August
Dec
December
Feb
Jan/Feb
January
June
Mar/Apr
March/April
May-June
May/June
Nov
Nov/Dec.
November
November/December
October
Sept / Oct
Sept/Oct
September

Entity Relationship Diagram Mock-up

Currently, FBN data is being kept in a flat excel file. Eventually, we would like to keep this information in a relational database using tables. This will allow the information to avoid repetition be more easily managed, modified, and searched. To the right is a mock-up of a potential table layout we could use in a relational database.



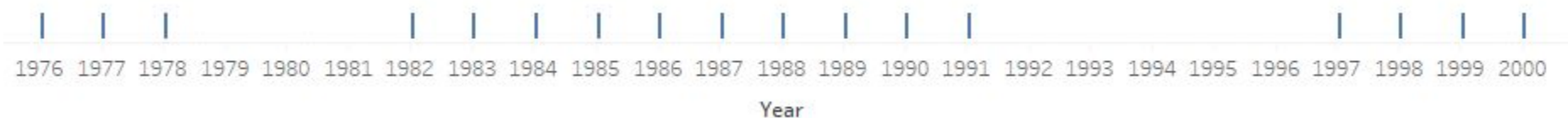
Trends Over Time





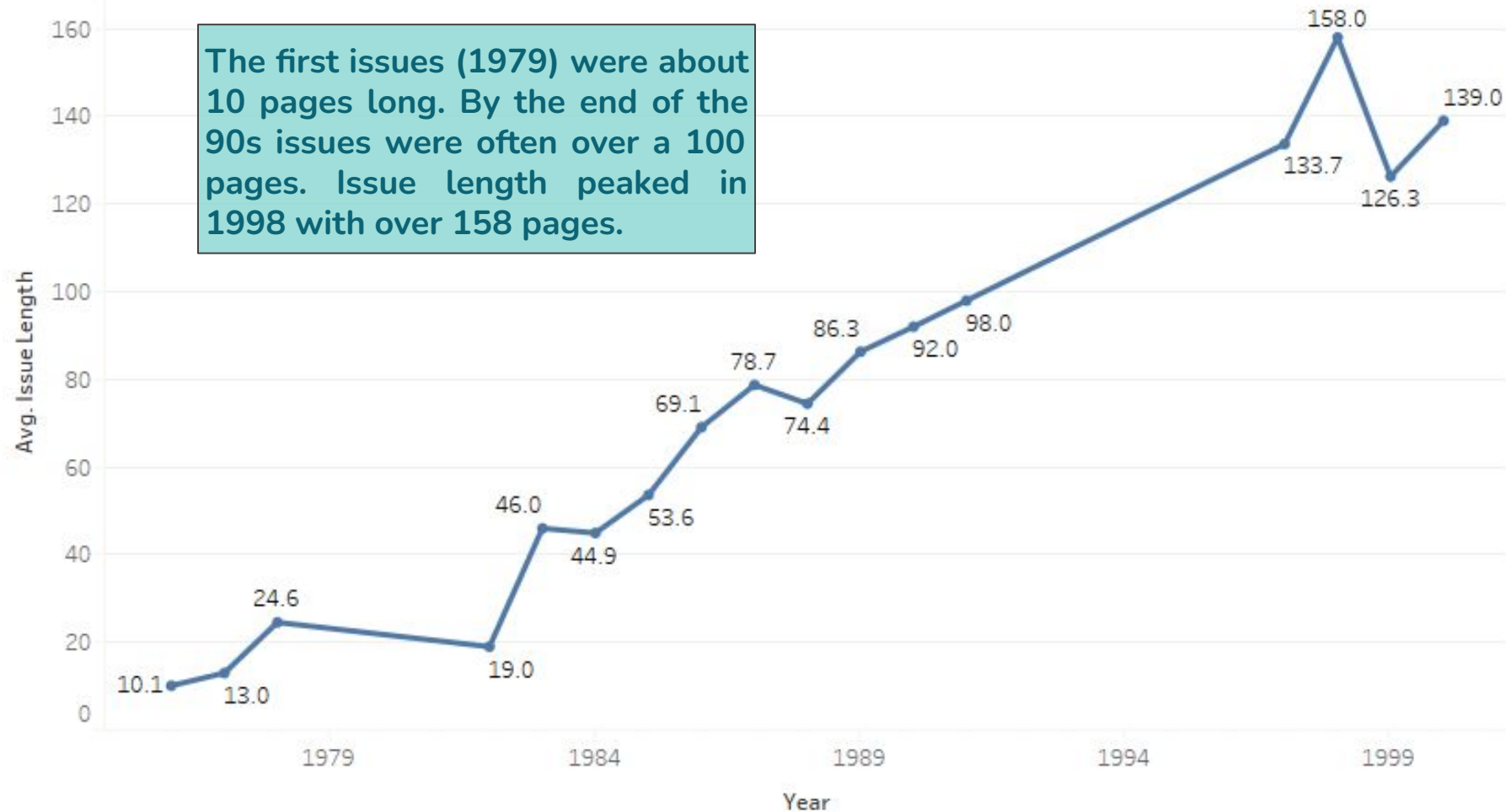
Missing Data

Our goal is to uncover trends over time. Preliminary data analysis is complete. It is important to note that we do not have data available for every year the newsletter ran. The image below shows which years are still missing. Available years have a tick and missing years do not.

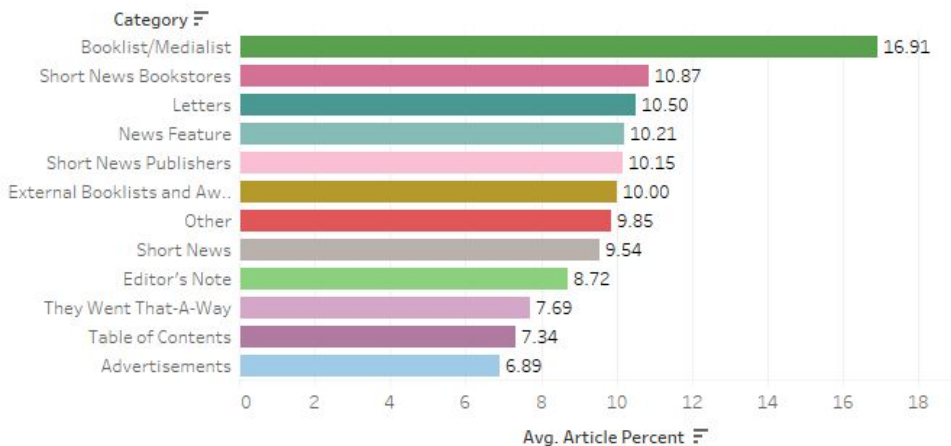


We are missing data for 1979, 1980, 1981, 1992, 1993, 1994, 1995, and 1996.

Issue Length Over Time



Category Percent in Issues (1970s)



1970s

- Booklist/Medialist was ~17% of issue content
- Advertisements were ~7% of issue content

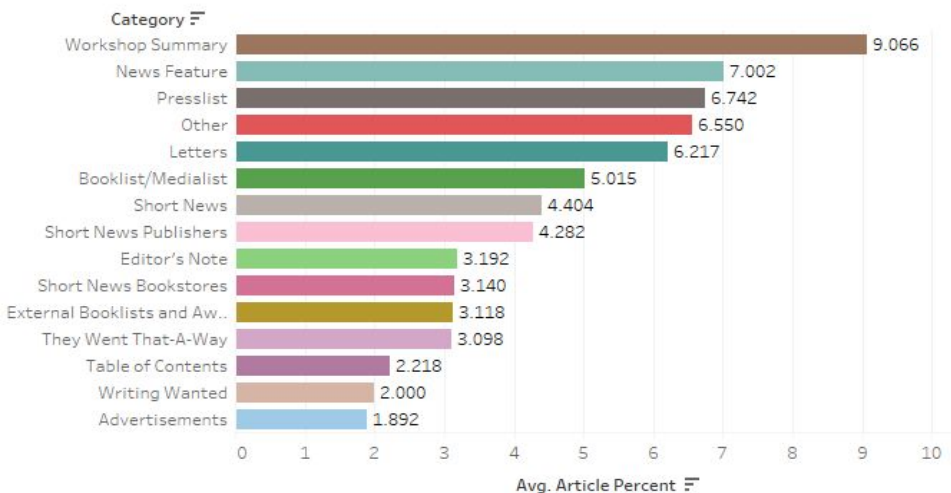
1980s

- Workshop Summary was ~9% of issue content
- Advertisement was ~2% of issue content

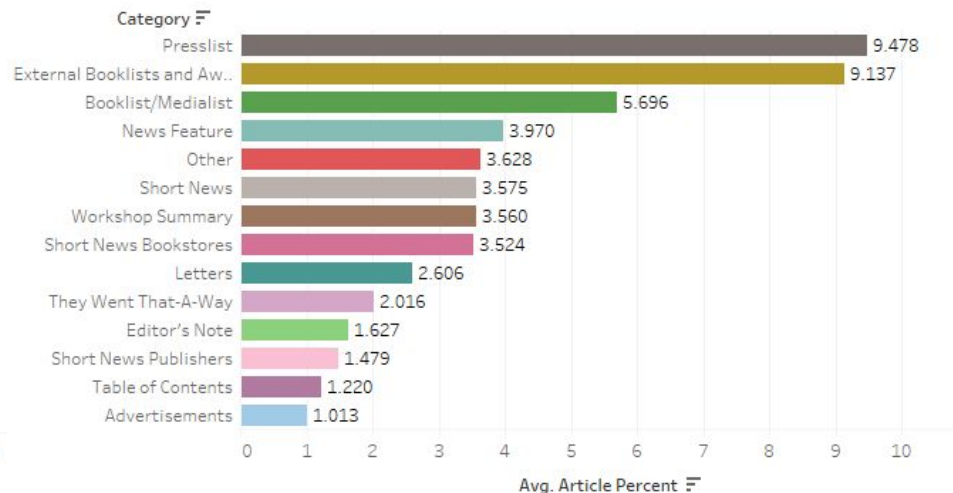
1990s

- Presslist was ~10% of issue content
- Advertisements were ~2% of issue content

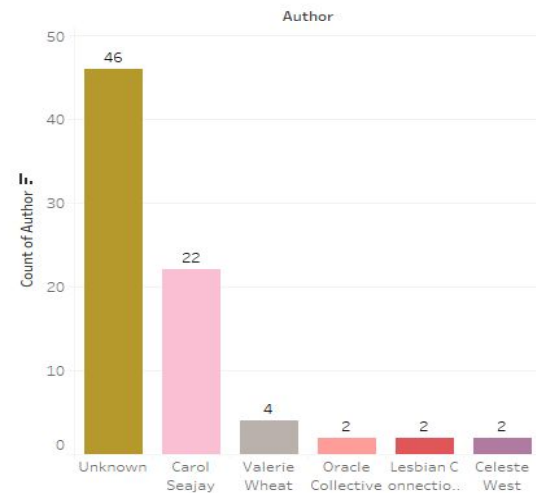
Category Percent in Issues (1980s)



Category Percent in Issues (1990s)



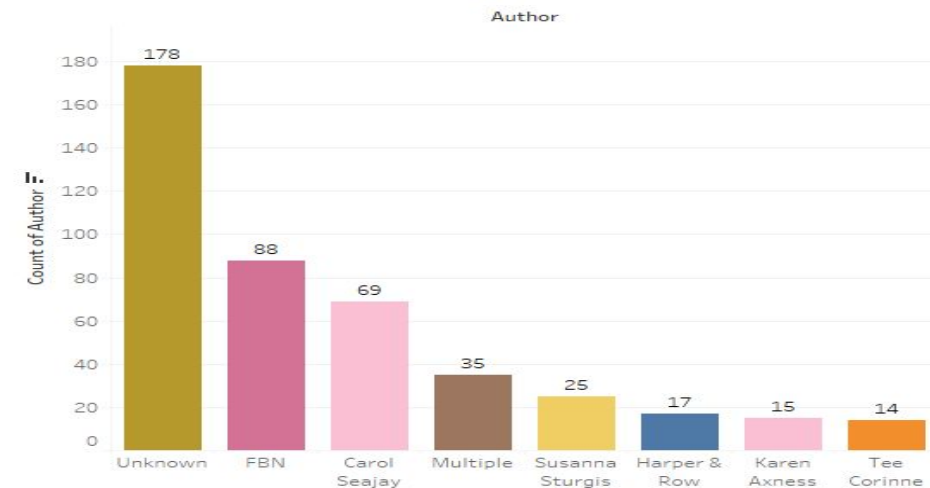
Author Count (1970s)



Carol Seajay has been the most prolific author. We can see the diversity of authors increases over time. 'Unknown' does take up a significant portion of the data. As we continue this project we should consider how to handle unknown authors.

To view an animation of this information [click here](#). The play button can be found at the lower right corner of the screen.

Author Count (1980s)



Author Count (1990s)



Closing Ideas





What comes next?

- Continue considering trends we can visualize in the data
- Collecting more data by completing issues
- Scraping Issues
- Moving data to a relational database



Resources

The Manga Guide to Databases is an Excellent resource to learn more about database management.

Takahashi M, Azuma S. Databases
Databases Databases Th E M a N G a G U
I de TM To. Accessed August 23, 2023.
<https://oberstar.eu.org/share/Documents/The-Manga-guide-to-databases.pdf>

Data Access

Below are links to my Tableau Public Access and Github where I am storing data related to the project

<https://github.com/SSGrasland/FBN->

<https://public.tableau.com/app/profile/salome.grasland>