Data Discovery: Examining Real-World Data for Future Database Classes

Omar Al-Jaafari & William Weinberger
Advisor: Dr. Kathryn Lenth

WESTMINSTER COLLEGE

QUESTIONS

What data is out there that can be taught to a databases class?

How can we find the proper data?

What sources have quality data?

How can this data teach the Structured Query Language (SQL) programming language and relational database concepts?

METHODS

Dataset Research

- Google Search, for finding datasets from the internet
- Kaggle, for easily finding well made datasets, despite them being made primarily for Machine Learning

Technologies

- MySQL, for hosting and setting up the databases [14]
- Python, for importing the data into the databases [13]
- Google Sheets, for data exploration and storing files in the cloud
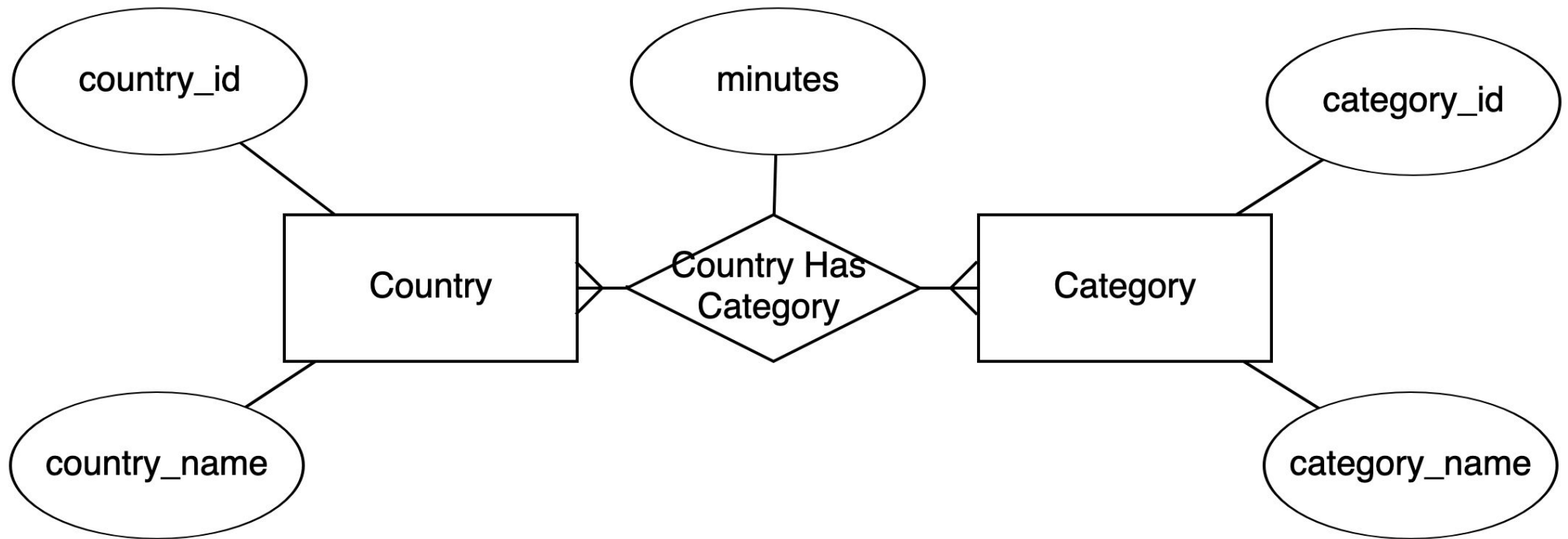- Excel, for data exploration

FIGURES

Figure 1a: Diagram of how much time is spent each day per country on average, and how that is stored [3], [4]. It also shows how tables (rectangles) are connected to each other and the columns (circles) they have. The diamonds represent relationships between tables.
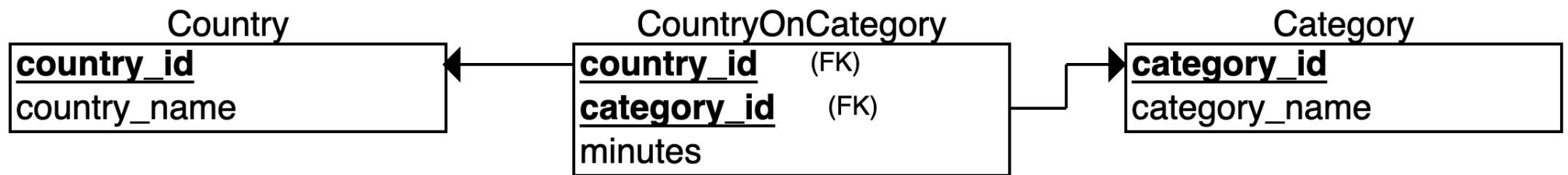
Figure 1b: Diagram of how much time is spent each day per country on average, and how that is stored [3], [4]. Tables can have arrows, which signify how one row labeled with a Foreign Key (FK) represents the row it points to.
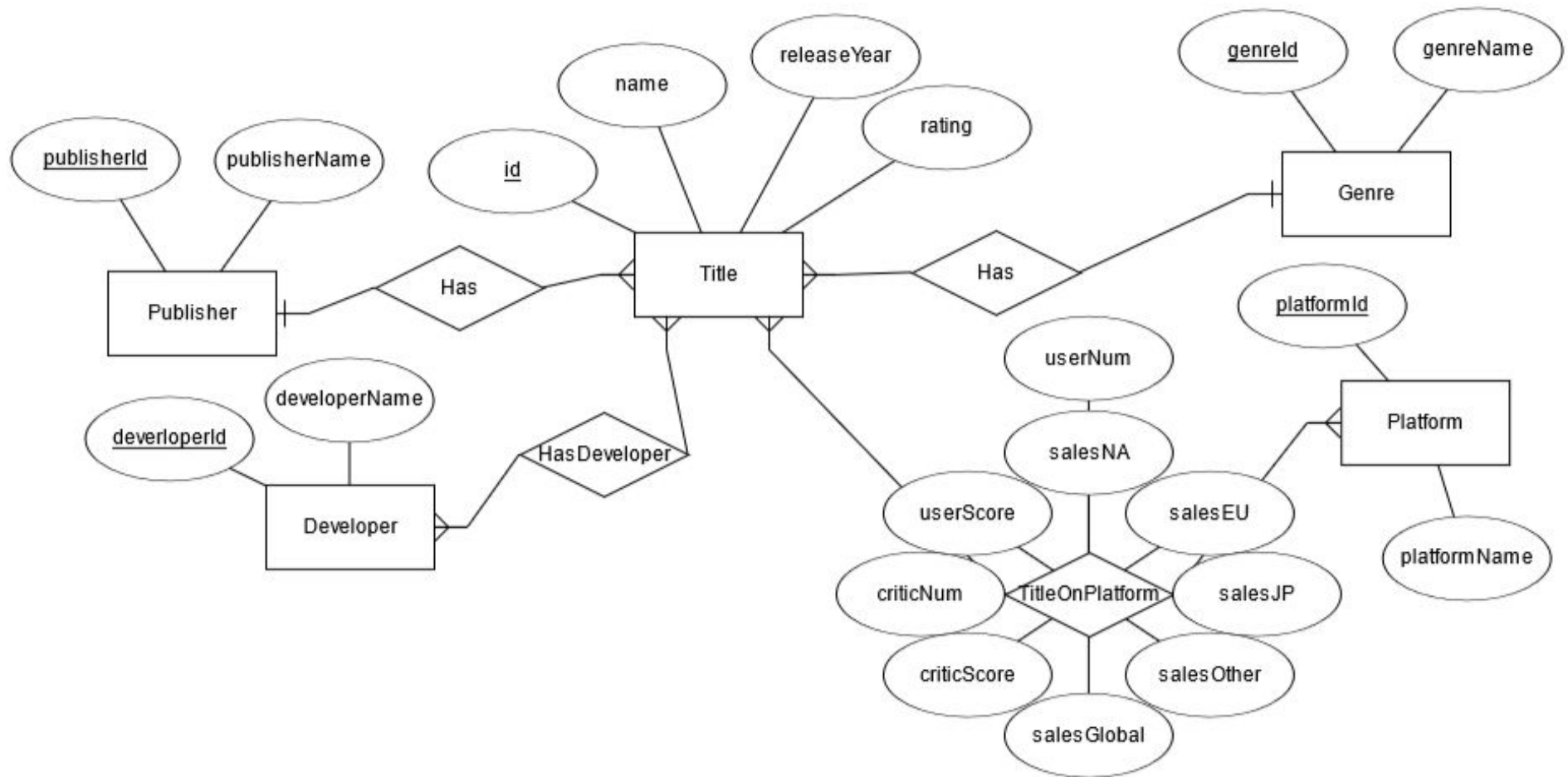
Figure 2a: Diagram of video game sales and score information [2], [10], [11]. It shows how tables (rectangles) are connected to each other and the columns (circles) they have. The diamonds represent relationships between tables.

Title
<u>**id**</u>
name
releaseYear
rating
genreId   (FK)
publisherId   (FK)

Has
<u>**id**</u>   (FK)
<u>**deverloperId**</u>   (FK)

Genre
<u>**genreId**</u>
genreName

Developer
<u>**deverloperId**</u>
developerName

Publisher
<u>**publisherId**</u>
publisherName

Platform
<u>**platformId**</u>
platformName

TitleOnPlatform
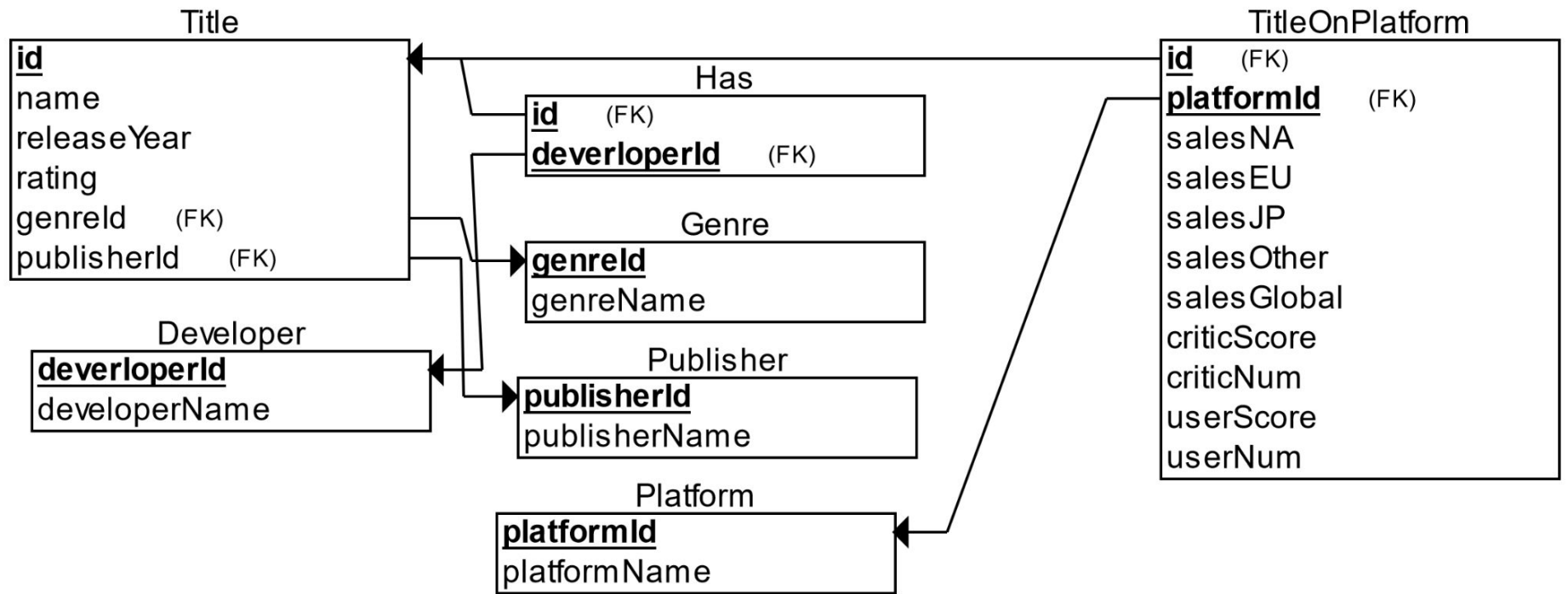| |
|---|
| <u>**id**</u>   (FK) |
| <u>**platformId**</u>   (FK) |
| salesNA |
| salesEU |
| salesJP |
| salesOther |
| salesGlobal |
| criticScore |
| criticNum |
| userScore |
| userNum |

Figure 2b: Diagram of how video game information is stored [2], [10], [11]. Tables can have arrows, which signify how one row labeled with a Foreign Key (FK) represents the row it points to.

DISCUSSION

- Vast amounts of data

- Decisions were made on database and data organization

- Code in Python was written to import data into tables in the database

FUTURE DIRECTION

- Big Data is vital for modern technology (social media, machine learning)
- Databases are useful for holding vast quantities of data for a variety of purposes
- Databases can help people solve problems (scientific research, business solutions)
- Databases we built will be used to help future students learn how to manage databases of all sizes and types

REFERENCES

[1]   National Park Service. "Biodiversity in National Parks." Kaggle. https://www.kaggle.com/datasets/nationalparkservice/park-biodiversity (accessed Jun. 23, 2022)

[2]   R. Kirubi. "Video Game Sales with Ratings." Kaggle. https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings (accessed Jun. 13, 2022)

[3]   E. Ortiz-Ospina, C. Giattino, and M. Roser. "Time Use." OurWorldInData.org. https://ourworldindata.org/time-use (accessed Jun. 15, 2022).

[4]   Organisation for Economic Co-Operation and Development. "Time Use." Organisation for Economic Co-Operation and Development. https://stats.oecd.org/Index.aspx?DataSetCode=TIME_USE (accessed Jun. 15, 2022)

[5]   R. Keizer. "International_Report_Departures." U.S. Department of Transportation. https://data.transportation.gov/Aviation/International_Report_Departures/innc-gbgc (accessed Jul. 15, 2022).

[6]   D. Morgan. "Database Profile: Aviation Support Tables." Bureau of Transportation Statistics. https://transtats.bts.gov/DatabaseInfo.asp?QO_VQ=IMI&DB_URL=Z1qr_VQ=E&Z1qr_Qr5p=N8vn6v10&f7owrp6_VQF=D (accessed Jul. 15, 2022).

[7]   United Nations Statistics Division. "Explorer." UNdata. http://data.un.org/Explorer (accessed Jul. 18, 2022).

[8]   Kaggle. "Learn Python, Data Viz, Pandas & More." Kaggle. https://www.kaggle.com/learn (accessed Jun. 13, 2022).

[9]   National Park Service. "NPSpecies." National Park Service. https://irma.nps.gov/NPSpecies/ (accessed Jun. 23, 2022)

[10]   G. Smith. "Video Game Sales." Kaggle. https://www.kaggle.com/datasets/gregorut/videogamesales (accessed Jun. 13, 2022)

[11]   VGChartz. "Video Game Charts, Game Sales, Top Sellers, Game Data." VGChartz. https://www.vgchartz.com (accessed Jun. 13, 2022).

[12]   Metacritic. "All Video Game Releases." Metacritic. https://www.metacritic.com/browse/games/release-date/available (accessed Jun. 13, 2022).

DOCUMENTATION

[13]   Python Software Foundation. "Python Docs." Python. https://docs.python.org/3/ (accessed Jun. 13, 2022).

[14]   Oracle. "MySQL Documentation." MySQL. https://dev.mysql.com/doc/ (accessed Jun. 7, 2022).

ACKNOWLEDGEMENTS

Dr. Kathryn Lenth

Dr. Krista Todd

Westminster College Summer Research Program

Family and friends