

METADATA - GENDER REPRESENTATION IN VIDEO GAMES

November 2022

Dataset by Brisa Palomar

INTRODUCTION

This dataset contains information on several video games released between 2012 and 2022. The games were chosen based on a series of criteria detailed later in this document.

The goal of this dataset is to compile information on the games and their characters to analyze how genders are represented.

GAME SELECTION CRITERIA

1. The games must have a storyline. To analyze the paper of each character is essential that games had a plot where characters had a role assigned (even if this role might change based on the player's choices). This excludes games like:
 - Puzzle games: Tetris, Candy Crush, Minesweepers, etc.
 - Racing games: Gran Turismo, Formula 1, Mario Kart, etc.
 - Social Simulators: Animal Crossing, The Sims, etc.
 - MMORPGs, where the storyline lived by the player, might considerably differ from other players, such as World of Warcraft.
 - Shooters with no story mode: Fornite, Valorant
 - Other popular games with no storyline: Minecraft, Roblox, League of Legends...
2. For games that offer a story and multiplayer modes (like some Call of Duty, GTA V...), just the story mode is taken into consideration for this analysis.
3. The games were selected for being top-selling or best-rated games of the year.
4. At least 5 games were selected for each year.

DATA SOURCE

This data has been compiled through research of several information sources including, but not limited to:

- Websites and media with a strong focus on video games like Metacritic, Destructoid, IGN, or GameSpot.
- Wikipedia
- The websites of the games developers, games publishers, and the website of the game itself.

ADDING TO DE DATASET

If anyone wants to collaborate with this dataset by adding or modifying information don't hesitate to contact me at [brisadataanalytics\(at\)gmail.com](mailto:brisadataanalytics(at)gmail.com) and we can discuss any changes or if you want to add information about different videogames I will email you a template to do so. Thanks!

DATA

The data is divided into three different data frames.

The table 'Games' contains data on the games; each row is a different game. To avoid bias in the game rating, 4 different reviewers' punctuations were collected to calculate an average.

The table 'Characters' contain information on the characters that are relevant to the story of the game.

The table 'Sexualization' is a rating based on 4 criteria to determine if a character is sexualized or not.

The tables relate as follows:

Games.Game_Id = Characters.Game

Characters.Id = Sexualization.Id

Games

Game_Id (str) - Primary Key - A unique set of letters and numbers that identifies a game.

Title (str) - the title of the game

Release (Date) - The date when the game was first released.

Series (str) - Series where the game belongs, if any.

Genre (str) - The main genre of the game

Subgenre (str) - Main subgenre of the game

Developer (str) - Game developer

Publisher (str) - Game publisher

Country (str) - Country of the game developer

Platform (str) - Platforms where the game is available. In the case of the game being available on other platforms years later, just the original is noted.

PEGI (int) - the Pan-European Game Information rating for that game. It indicated the minimum age recommendation for a video game.

Customization (str) - If the game offers the option of customizing or not the character. It contains three values: 'Yes', 'No', and 'Non-Binary'.

Protagonist (int) - number of protagonists for that game.

Protagonist_non_male (int) - number of non-male protagonists. This is, females, non-binaries, and customizable characters.

Relevant_males (int) - male characters in the game.

Relevant_no_males (int) - non-male characters in the game

Percentage_non_male (float) - the percentage of non-male characters

Criteria (str) - criteria for selecting the game. Contains three values:

- 'TR' - top rated
- 'MS' - most sold
- 'SR' - sales and rating.

Director (str) - game director gender. Can contain 4 values:

- 'M' - male
- 'F' - female
- 'NB' - non-binary
- 'B' - both in case there is more than one director of different genders.

Total_team (int) - number of main people involved in the game creation. Includes main programmers, developers, directors, producers, artists, and designers.

Female_team (int) - number of team integrants that are female.

Team_percentage (float) - the percentage of women in the team.

Metacritic (float) - punctuation out of ten given to the game by Metacritic.

Destructoid (float) - punctuation out of ten given to the game by Destructoid.

IGN (float) - punctuation out of ten given to the game by IGN.

GameSpot (float) - punctuation out of ten given to the game by GameSpot.

Avg_reviews (float) - the average of the four previous columns.

Characters

Name (str) - the name of the character

Gender (str) - the gender of the character. It contains 4 different values:

- 'Female' - characters identified as females in the game. They are addressed with the pronouns she/her.
- 'Male' - characters identified as males in the game. They are addressed with the pronouns he/him.

- 'Non-binary' - characters whose gender is purposely left ambiguous, those that due to their nature don't have a gender and no gender has been assigned to them (animated object, plants, animals...), and those who self-identify as non-binaries. They are addressed using the pronouns they/them.
- 'Custom' - those characters that the game offers the option to customize their gender. They are addressed according to the gender chosen by the player.

Game (str) - foreign key - the Game_Id from the 'Games' data frame according to the character's game.

Age (str) - the age of the character during the game events.

Age_range (str) - a ranking categorization of the ages. The values are as follows:

- Infant - 0 to 5 years old
- Child - 6 to 14 years old
- Teenager - 15 to 17 years old
- Young-Adult - 18 to 24 years old
- Adult - 25 to 39 years old
- Middle-Aged - 40 to 64 years old
- Elderly - older than 65 years old
- Unknown - characters whose age is unknown

Playable (boolean) - if the character is playable. It can be either:

- 0 - No
- 1 - Yes

Sexualization (int) - a punctuation out of 4 determined by the data frame 'Sexualization'.

Id (str) - Primary Key - an id unique for each character.

Species (str) - the species of the character.

- Human - all those characters with human-like appearances. This includes humans but also other fantasy races whose appearance is basically that of a human with very little physical divergences like gods, elves, dwarfs...
- Humanoid - characters whose appearance is anthropomorphic but their physical appearance makes it obvious that they are not human. Animal-Humanoids do not belong here as they have their own category.
- Humanoid-Animal - characters that are a combination between a human and a real-life animal. The animal features must be recognizable enough to know the type of animal, otherwise, they are considered Humanoids.
- Android-Robot - androids, robots that have a human appearance.
- Robots - robots that do not have a human appearance. This includes artificial intelligences without a physical body.
- Animated object - characters whose physical appearance is that of an object, but that has been given consciousness and some human features such as a face, arms or legs.
- Animated plant - plants that keep their plant appearance but that have been given some human features such as a face, arms, or legs.
- Animal - characters that are physically recognizable real-life animals. They might have some special features such as speech ability or magic powers, but it does not affect their looks.
- Creatures - all those creatures that do not fit in any of the above categories. It includes monsters whose physical appearance is too distant from that of a human to be considered humanoids, fantastic animals that do not look like any real-life animal, mythological creatures...
- Unknown: those characters' whole physical appearance is never revealed in the game.

Side (str) - the side in which a character is.

- P - on the protagonist's side. This includes characters that are supportive or neutral towards the protagonist.
- A - antagonists
- B - characters that are or can be on both of the above sides.

Relevance (str) - the relevance of a character in a game and in relation to the protagonist

- PA - protagonist - the most important person(s) in the game. In case of is more than one both of them must have the same importance in the plot.
- DA - deuteragonist. The second most important character(s) in the plot. In case of being more than one, both must have the same relevance.
- SK - sidekick. Those characters accompany the protagonist during all, or most of, the story. They offer constant support to the protagonist by giving advice, battle support, exploration aid, etc. They differ from the deuteragonist as they usually have little to no relevance in the storyline.
- MC - the main character. A character that is relevant throughout all, or most part of, the story. This category can include antagonists.
- SC - secondary character. A character that is important in the storyline but whose relevance is occasional, be it because their plot just lasts a short amount of time or because they are mentioned throughout the game but barely appear in-game. This category can include antagonists.
- MA - main antagonist. The main antagonist(s) of the game. It is relevant throughout the game.

Romantic interest (str) - If that character is, or can be, the romantic interest of the protagonist. This includes one-sided romances as long as the one who holds the romantic feelings is the protagonist and also love-less relationships as long as the characters are married. A protagonist only can be a romantic interest if there is more than one protagonist and they are each other's romantic interest.

- No - this character is no romantic interest of the protagonist
- Yes - this character is unavoidably the romantic interest or partner of the main character.
- Opt - this character can be dated or can have romantic or sexual interactions with the protagonist based on the player's choice.

Sexualization

Id (str) - Primary Key - the id identifying the character evaluated in this table.

Sexualized clothing (boolean) - if the clothes of a character are sexualized (1) or not (0). They just need to meet one of these criteria to be TRUE:

- The character is wearing clothes that are not age-appropriate: too sexualized adult clothes on children or children-like sexualized clothes on adults.
- The character is wearing clothes that make no sense: high-heels for running and fighting, too revealing clothes in extreme weather conditions such as heavy rain or cold, and armor that does not cover vital organs.
- Just the female character(s) wears clothes that reveal sexualized areas of the human body such as cleavage, upper tights, lower belly, or buttstock, while her male counterparts don't.

Trophy (boolean) - the character is a trophy for a main character (1) or not (0). The only purpose of the character in the game is to be 'won' as a prize for a male character. The character is objectified.

Damsel in distress (boolean) - the character has the role of a damsel in distress (1) or not (0). The character in-game exists to be rescued or saved by the protagonist or the main character. If the character has this role but their age is that of a child or an infant this category is considered as negative, as a child needing the protection or help of an adult is not considered a sexist role.

Sexualized cutscenes (boolean) - some or all the cutscenes where this character appears are sexualized by portraying them in a suggestive way or unnecessarily focusing on areas of the body that are usually considered sexually appealing such as cleavage, breasts, pubic area, and buttstock.