Data Appendix

Predicting a Player's Stardew Valley Marriage Candidate

Sydney Weisberg, Kika Kovaleski, Molly Daniel, Caroline Zouloumian, Emma Ruckle

2025-04-14

Part I: Loading and Cleaning the Data

```
library(tidyverse)

stardew_all <- read.csv("stardew_all.csv") |> select(!Timestamp) |> mutate(from =
↪  "All")
stardew_smith <- read.csv("stardew_smith.csv") |> select(!Timestamp) |>
↪  mutate(from = "Smith")

stardew_data <- rbind(stardew_all, stardew_smith)
```

Part II: Describe the Data

Dataset: Stardew Valley Dataset

- Filename: stardew_all.csv and stardew_smith.csv

- gender: gender of the survey respondent

  - Categorical data
  - Categories: Male, Female, Non-binary, Other

- sexuality: sexuality of the survey respondent

  - Categorical data
  - Categories: Lesbian, Gay, Bisexual, Straight, Other

- age: age of the survey respondent

  - Continuous numeric data
  - Range: 0-100

- relationship: relationship status of the survey respondent

  - Categorical data

- – Categories: Single, Dating, Married, Other

- hobbies: preferred hobby of the survey respondent

  - – Categorical data
  - – Categories: Sports, Arts, Music, Literature, Cooking

- activity: preferred Stardew activity of the survey respondent

  - – Categorical data
  - – Categories: Mining, Fishing, Decorating Farm, Exploring, Farming, Talking to NPCs, Combat

- activity: preferred Stardew season of the survey respondent

  - – Categorical data
  - – Categories: Spring, Summer, Fall, Winter

- activity: preferred Stardew season of the survey respondent

  - – Categorical data
  - – Categories: Spring, Summer, Fall, Winter

- personality: personality type of the survey respondent

  - – Categorical data
  - – Categories: Introvert, Extrovert, Ambivert

- playtime: times the respondent has played Stardew through and romanced a character

  - – Continuous numeric data
  - – Range: 0-200

- non_marriage_candidate: favorite non-marriage candidate of the survey respondent

  - – Categorical data
  - – Categories: Caroline, Clint, Demetrius, Dwarf, Evelyn, George, Gus, Jas, Jodi, Kent, Leo, Lewis, Linus, Marnie, Pam, Pierre, Robin, Sandy, Vincent, Willy, Wizard

- marriage_candidate: favorite marriage candidate (or Krobus) of the survey respondent

  - – Categorical data
  - – Categories: Abigail, Alex, Elliot, Emily, Harvey, Haley, Krobus, Leah, Maru, Penny, Sam, Sebastian, Shane

Part III: Pressing Issues

1. We have two subsets of our data - one from Smith College and one largely from Stardew Valley Reddit and Discord servers. The first one is biased due to the sample coming from Smith, and the second one is biased due to it mainly consisting of more hard-core Stardew Valley players. We have combined the datasets together, which influences the bias. Since there is more data from the non-Smith group, the overall data is more influenced by that group.
2. As a whole, our sample size is relatively small. Within our data, some romanceable candidates were only chosen by one person. Since it is small, it likely does not reflect the actual population of Stardew Valley players.
3. Since we chose both the questions and the responses to ask, we are inherently biasing the infrastructure of our data set. Not everyone will fit into some of the categories we listed, and thus we are limiting self-expression.