

Table of contents

1	Demographic, Behavioral, & Quiz Surveys	2
1.1	Main Demographic Surveys	2
1.2	Dataset Quick Profile	3
1.3	Major Variable Groupings	3
1.4	Quiz Responses (Master)	4

1 Demographic, Behavioral, & Quiz Surveys

This section describes the individual-level demographic and behavioral survey datasets, as well as the specialized quiz responses. You can download the complete, de-identified longitudinal panels below.

1.1 Main Demographic Surveys

[Download CSV Dataset \(.csv\)](#) [Download R Dataset \(.rds\)](#) [Download Stata Dataset \(.dta\)](#)
[Download SPSS Dataset \(.sav\)](#)

Interactive Survey Instruments

Explore the fully digitized survey instruments for each wave below:

- [Wave 1: Summer 2011 Demographic Survey](#)
- [Wave 2: Spring 2012 Demographic Survey](#)
- [Wave 3: Summer 2012 Demographic Survey](#)
- [Wave 4: Fall 2012 Demographic Survey](#)
- [Wave 5: Winter 2013 Demographic Survey](#)
- [Wave 6: Summer 2013 Demographic Survey](#)

This dataset combines survey responses from 203 Notre Dame students across 6 consecutive waves, capturing demographic backgrounds, personality inventories (e.g., Big Five), academic metrics (GPAs), musical and artistic preferences, lifestyle behaviors, and campus engagement.

! Security & Privacy Scrubbing

The raw demographic dataset originally contained sensitive participant identifiers. A secure, programmatic R script (`Code/scrub_demographics.R`) has been run to drop 38 personal identifying columns containing: - Participant first names and last names (`firstname_*`, `lastname_*`, `v1532`, `v1180`, etc.) - University NetIDs and emails (`netid`, `netid_*`, `emailaddress_*`) - Raw cellular telephone numbers (`phonenum`)

1.2 Dataset Quick Profile

Below is a quick programmatic summary of the clean demographic survey data loaded directly from the processed CSV.

```
library(readr)
library(dplyr)

# Load clean demographics dataset
dem_df <- read_csv("Surveys/demographics_longitudinal_clean.csv", show_col_types = FALSE)

# Profile metrics
profile <- tibble(
  Metric = c("Total Students (Egos)", "Total Survey Attributes", "Sensitive Columns Safely Dropped"),
  Value = c(nrow(dem_df), ncol(dem_df), 38)
)

knitr::kable(profile, caption = "Processed Demographic Survey Profile")
```

Table 1.1: Processed Demographic Survey Profile

Metric	Value
Total Students (Egos)	203
Total Survey Attributes	1556
Sensitive Columns Safely Dropped	38

1.3 Major Variable Groupings

With 1,556 attributes, variables are suffixed with a wave number (e.g. `_1` for Wave 1, `_6` for Wave 6) to track changes over time. Some of the core groupings include:

- **Demographics:** Participant gender (`gender_1`), ethnicity, family background, and dorm locations.
- **Academic Performance:** Semester-by-semester GPAs (e.g., `firstsemgpa_2`, `gpa_6`).
- **Personality Scales:** Big Five traits (Extraversion, Neuroticism, Openness, Agreeableness, Conscientiousness).
- **Lifestyles and Habits:** Alcohol use, smoking, sleep habits, and smartphone usage metrics (e.g., `phoneuse*_1`).

- **Cultural Tastes:** Music genres, movies, books, and artistic preferences (e.g., `bigband_1`, `classicrock_1`, `jazz_1`, `rap_1`).
-

1.4 Quiz Responses (Master)

In addition to the main demographic surveys, the study administered a series of quizzes to participants. These quizzes capture high-frequency assessments of student health, emotional well-being, political discussions, and feedback on their smartphone performance.

[Download Quiz CSV \(.csv\)](#) [Download Quiz Stata \(.dta\)](#)

PII Scrubbing

The raw quiz datasets originally tracked participants using raw 10-digit phone numbers. These have been securely mapped to the standard `egoid` identifier and the raw phone numbers have been completely dropped to maintain anonymity.

The consolidated quiz dataset contains responses from 197 students across 146 specialized survey items, all of which are documented in the [Searchable Variable Catalog](#).