Proposed Revision- OkCupid Data for Introductory Statistics and Data Science Courses

Albert Kim and Adriana Escobedo-Land

February 10, 2020

Goal of Changes

- Remove personally identifiable data from essay data and provide new dataset

- Maintain componenets of paper for research and pedagological value

- Create data ethics example of the work

Suggested Actions

The Manuscript

- Pg 3 — Section 2 — Paragraph 1 — Line 2: Remove month and day from date "date redacted"

- Pg 4 — Section 3 — Paragraph 1 — Line 4: Remove question 3; add "essay data redacted"

- Pg 8 — Section 3.3.1 and 3.3.2: Remove; add "essay data redacted"

The Data

- Remove metadata including month and date of data source

- Add noise to "age" field

- Remove "location" field

- Remove "last online" field

- Edit codebook accordingly