PRISM Quick Reference: Importing Fixed Width Data

Dino Christenson & Scott Powell

Ohio State University

February 15, 2008

Data Importing

- General data files (.dat or .txt) are usually presented in one of two formats: Delimited or Fixed Width

- If a file is delimited, this means that the values within the file are separated by a common "delimiter" (*e.g.* comma, tab, semicolon, etc.)

- If a file uses fixed widths, this means that each variable has a preset, consistent column size throughout the data file

The Problem

- Importing fixed width data into your preferred statistical software program

- How to know if this is your problem:
  - Your data is in a .dat or .txt file
  - It does not seem to be delimited in any obvious manner
  - A codebook or accompanying document gives you information on the column widths of each variable in the dataset
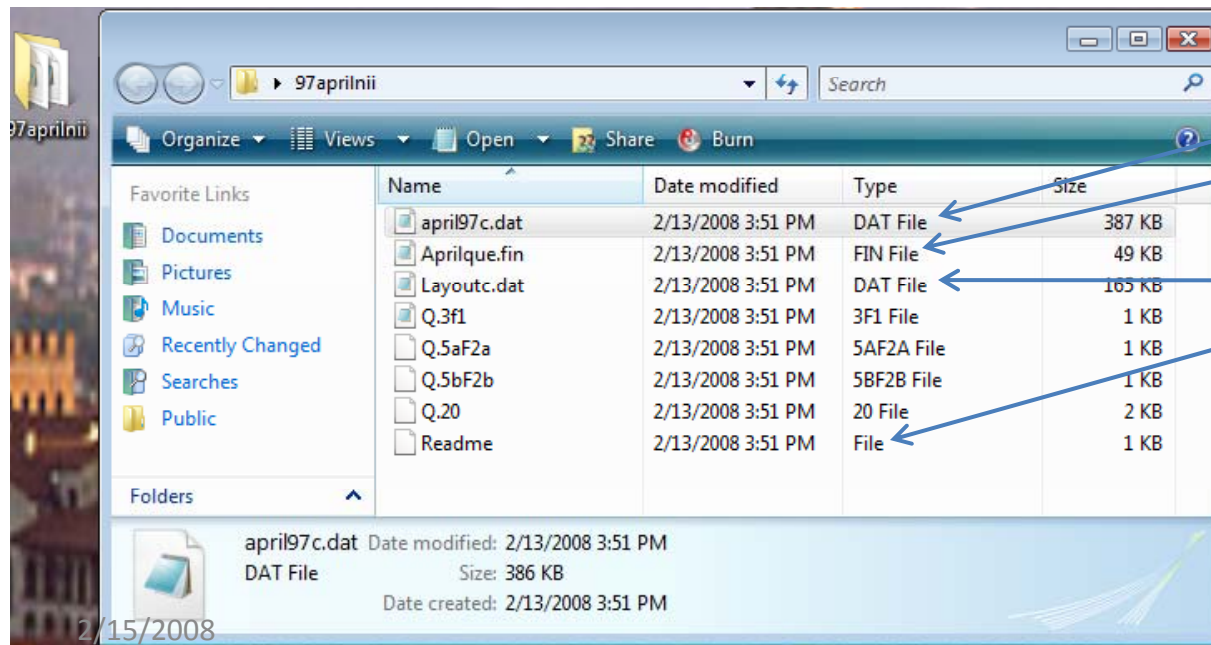
An Example: News Interest Index

- Say you wanted to know more about the public's news interest…
  - Pew has many related surveys
  - http://people-press.org/dataarchive/
- Let's download the April survey dataset and accompanying files

Identifying the Files

- A zip file is downloaded
  - Extract all the contents of the zip file into a new folder
- Open the new folder
  - Inside we find several files

Dataset (.dat)

Questionnaire (Word Perfect .fin)

Card Layout (.dat)

Read me (.txt)

For more on file extensions see:
http://polisci.osu.edu/prism/ImportingData.htm

Fixed Width Data

- **What does the data look like?**
  - Short answer: a mess
- **So how do we know which numbers correspond to which variables?**
  - Answer: fixed width columns
- **How do we know how many columns per variable?**
  - Answer: codebook

Codebook Column Code

- Look at the codebook and pay attention to the lines that tell you how many columns there are per question
  - For example the first item:
  - *** RESPONDENT # CARD 1 COLS 1 - 6 ***
  - Tells us that the respondent number variable is made up of the first six columns
- Why does this matter?
  - Essentially the data is stored as a non-delimited matrix and we will have to tell it which columns belong to which variable

Layoutc.dat - Notepad

File   Edit   Format   View   Help

```
Questionnaire name: 7281                    04/04/97 - 1:09 PM              Page:   1

*** RESPONDENT #  CARD  1 COLS 1 - 6 ***

< Batch Number >

-- CARD  1 COLS.  7- 9 --

< Interview Date >

-- CARD  1 COLS. 10-15 --

*** QUESTION #  1 ***
*SCHULMAN, RONCA & BUCUVALAS, INC., 32ND STREET, N.Y., N.Y.
.7281 - NEWS INTEREST SURVEY
.APRIL, 1997
.

*** QUESTION #  2 ***
*SAMPLE READ-IN: REGION (1 DIGIT)
  GO TO Q. #  3 ====>  < 1 > *NORTHEAST
  GO TO Q. #  3 ====>  < 2 > *SOUTH
  GO TO Q. #  3 ====>  < 3 > *MIDWEST
  GO TO Q. #  3 ====>  < 4 > *WEST
  GO TO Q. #  3 ====>  < 5 > [05]###
-- CARD  1 COL. 36 --

*** QUESTION #  3 ***
   << CONDITIONAL ASSOCIATED WITH THIS QUESTION >>
    IF QUOTA GROUP #  1 CELL #  1 IS FULL       (CONDITIONAL # 49)
   AND Q#  2 EQ CODE(S) 1                       (CONDITIONAL # 50)
   THEN GO TO Q.#D24 ELSE GO TO Q.#  3.
   << CONDITIONAL ASSOCIATED WITH THIS QUESTION >>
    IF QUOTA GROUP #  1 CELL #  2 IS FULL       (CONDITIONAL # 51)
   AND Q#  2 EQ CODE(S) 2                       (CONDITIONAL # 52)
   THEN GO TO Q.#D24 ELSE GO TO Q.#  3.
   << CONDITIONAL ASSOCIATED WITH THIS QUESTION >>
    IF QUOTA GROUP #  1 CELL #  3 IS FULL       (CONDITIONAL # 53)
   AND Q#  2 EQ CODE(S) 3                       (CONDITIONAL # 54)
   THEN GO TO Q.#D24 ELSE GO TO Q.#  3.
   << CONDITIONAL ASSOCIATED WITH THIS QUESTION >>
    IF QUOTA GROUP #  1 CELL #  4 IS FULL       (CONDITIONAL # 55)
   AND Q#  2 EQ CODE(S) 4                       (CONDITIONAL # 56)
   THEN GO TO Q.#D24 ELSE GO TO Q.#  3.
Questionnaire name: 7281                    04/04/97 - 1:09 PM              Page:   2

*SAMPLE READ-IN: CATI STATE CODE (2 DIGITS)
  GO TO Q. #  4 ====>  < 1 > *Alabama
  GO TO Q. #  4 ====>  < 2 > *Alaska
  GO TO Q. #  4 ====>  < 3 > #hold
  GO TO Q. #  4 ====>  < 4 > *Arizona
  GO TO Q. #  4 ====>  < 5 > *Arkansas
  GO TO Q. #  4 ====>  < 6 > *California
  GO TO Q. #  4 ====>  < 7 > #hold
```
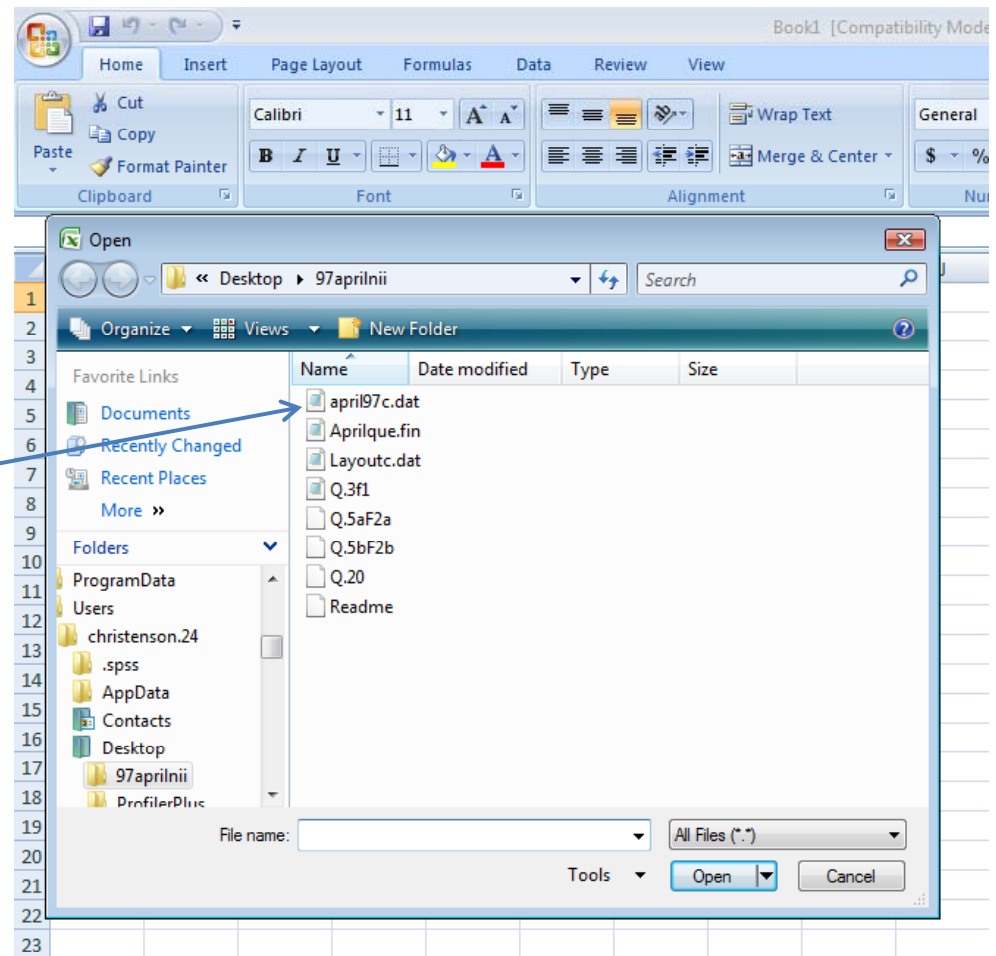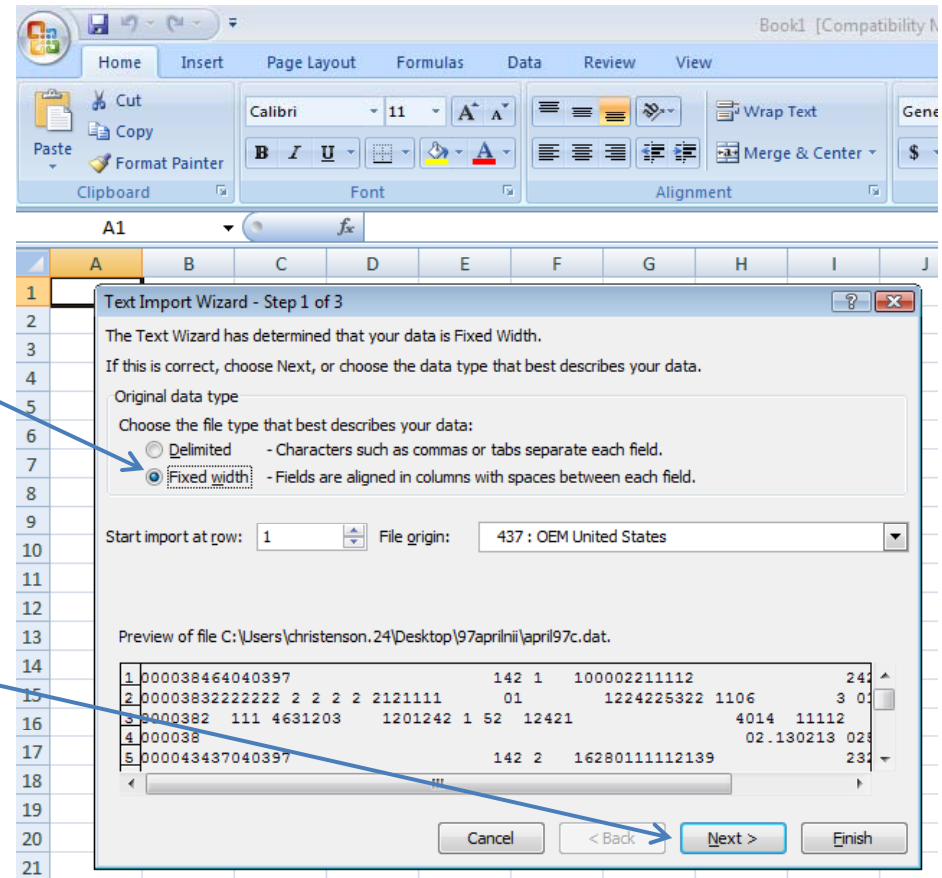
Importing to Excel (or SPSS)

- It's easiest to denote the fixed widths in SPSS or Excel
- After that its easy to transfer the data to your preferred program
- Let's try this in Excel (it's basically the same in SPSS)
  - Open up Excel
  - Select "Open"
  - Search for the dataset "april97c.dat"
- Select to open the dataset
  - One of two things will happen
  - A. Either you will be immediately prompted by the text import wizard OR
  - B. The data will get thrown in the excel sheet in the wrong order
    - In the case of B, click on the writing tablet GUI button in the corner of the data and select to use the Text Import Wizard
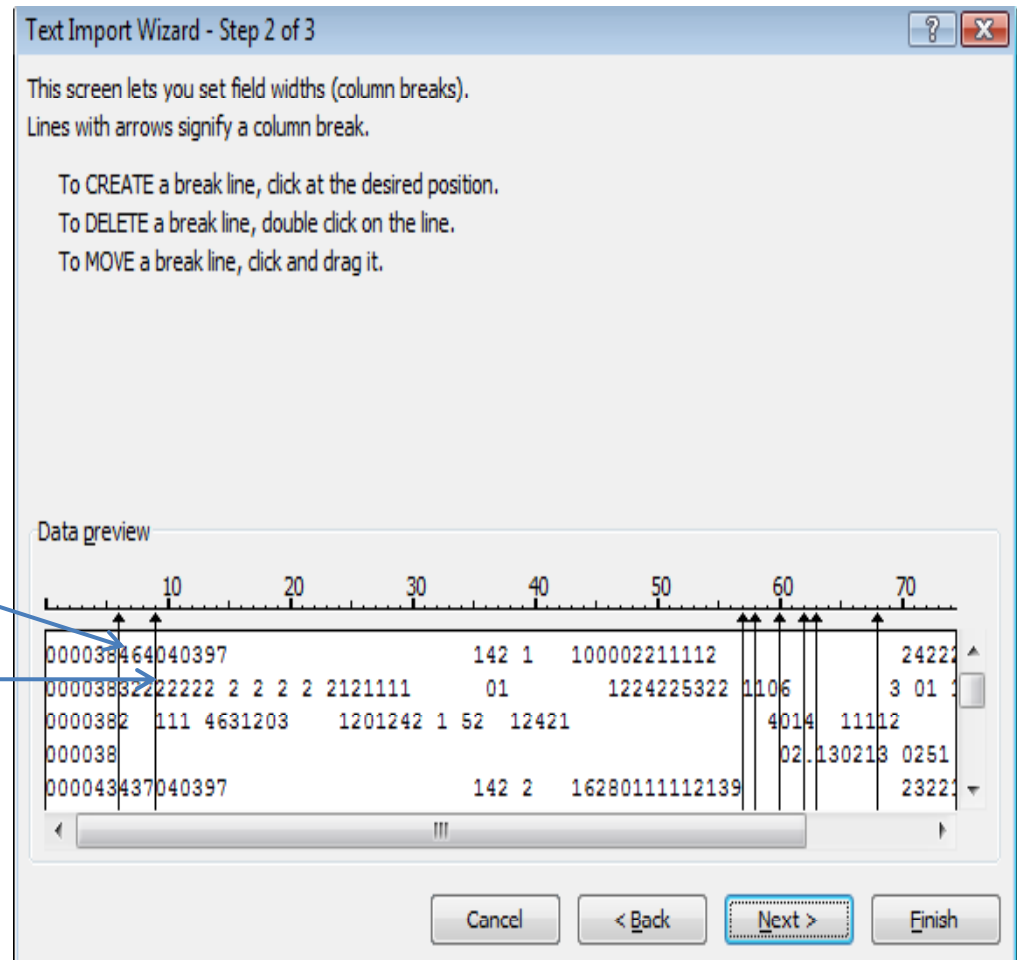
Setting Fixed Widths

- Once we are at the text import wizard
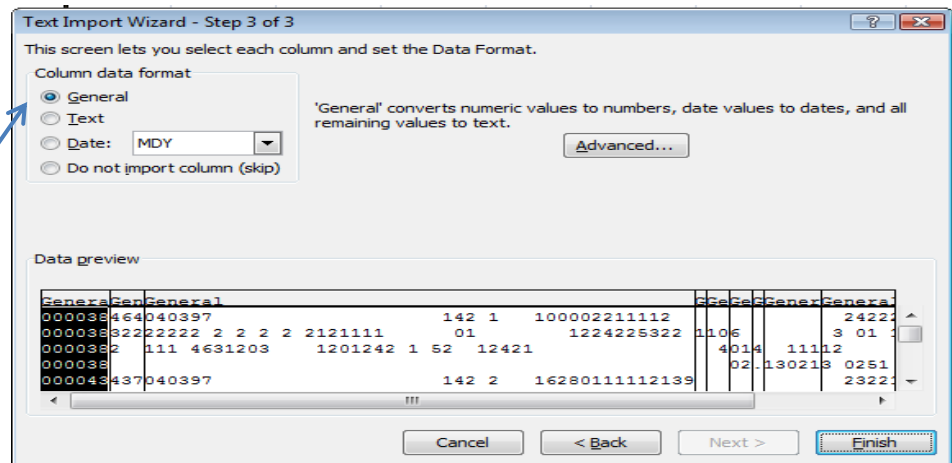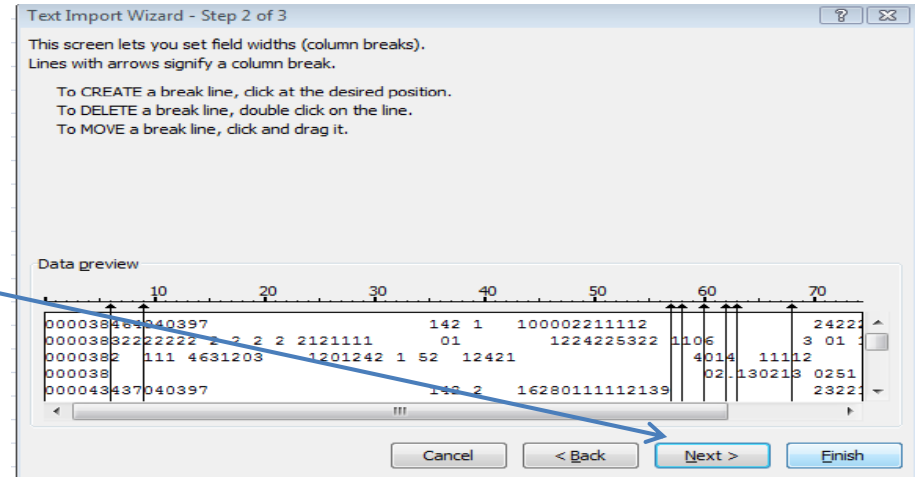- Select fixed width
- Select the next button

Setting Fixed Widths

- Here we can set the widths or column breaks for each variable
- Recall the first variable (slide 6):
  - *** RESPONDENT # CARD 1 COLS 1 - 6 ***
  - This tells us that the fist variable is made up of columns 1,2,3,4,5 & 6.
  - Therefore we place a break before column 7
- Then we refer back to the codebook for the second variable's fixed width, denote it with an arrow and so on until we have done this for all the variables in the set…
- Yes, it's tedious

Setting Fixed Widths

- After you've finished all the delimiting
  - Click next
- Now we can label the data format for each variable:
  - General, Text, Date...etc.
  - For e.g., we leave the first var as "general" because it is the respondent card
  - Then click Finish

Final Dataset

- Now you should have the appropriate dataset in excel format
  - Which is easily imported into various stats programs
  - For cross-software importing see our addendum http://polisci.osu.edu/prism/ImportingData.htm

  - Note: the dataset here is not completely formatted; i.e., as you can see, we only did the first two variables in this example, because it is too tedious to format all of them...

Good References

- [PRISM's Addendum on Data Importing](#)

- [PRISM's Introduction to Stata](#)

- [PRISM's Introduction to R](#)

- All available on the PRISM Brownbag site:
    - http://polisci.osu.edu/prism/luncheons.htm