Engineering or Hacking AIT A FEW WEEKS UNTIL IT BECOME OF VOICE OF "WANNA

Prof. Martin Shepperd

martin.shepperd@brunel.ac.uk

AND BECAUSE YOU'RE LUCKY PEOPLE AROUND YOU UNDERSTA

CREATE MORE PROBLEMS THA

MY CODE 15 SORT OF SIMILAR TO A

THAT YOU CAN EITHER COME LOOK

PollEv.com/mshepperd

Virtual seminar protocol

- 1. I will start at 1505 prompt; please be ready
- 2. Please mute your microphone
- 3. If you have a question can you use the chat facility or ...
- 4. ... or ask during in a question gap
- 5. ... or interrupt (it's fine)
- 6. Thumbs-up questions that are important to you
- 7. Nathan and Tasin are monitoring and answering chat thanks guys!
- 8. Be aware, the seminar will be recorded

Seminars and Labs

- I will lead the seminar
- Maybe easier to watch (and answer the Pollev.com questions)
- Then do the exercises in the gaps or lab afterwards
- I will leave the seminar and then join the lab meeting
- There are answers in the Worksheets

Seminar + Lab Agenda

- 1. FAIR data
- 2. A (bad) data sharing example
- 3. Debugging and getting help
- 4. Lab Exercises
 - User defined functions

1. FAIR Data¹

- Findability
- Accessibility
- Interoperability
- Reusability

¹ Wilkinson, Mark D, et al. 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." Scientific Data 3 (1): 1–9.

FAIR principles

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
- A1.1 the protocol is open, free, and universally implementable
- A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- 12. (meta)data use vocabularies that follow FAIR principles
- 13. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
- R1.1. (meta)data are released with a clear and accessible data usage license
- R1.2. (meta)data are associated with detailed provenance
- R1.3. (meta)data meet domain-relevant community standards

Source: Wilkinson et al., 2016

What is Meta-data?

- data about data
- meta-data describes how the data set is organised and the meanings of individual variables
- For example, we might say that the variable Person ID
 - comprises alpha-numeric characters (so it's a string)
 - to be valid it must be (i) unique and (ii) exactly 8 chars
- it facilitates meaningful processing and analysis

Meta-data

- F2: data are described with rich metadata
- A1: meta-data are retrievable by their identifier [DOI] using a standardized communications protocol
- I1: meta-data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- R1.2: meta-data are associated with detailed provenance
- R1.3: meta-data meet domain-relevant community standard

2. A (bad) data sharing example²

1208

IEEE TRANSACTIONS ON SOFTWARE ENGINEERING. VOL. 39. NO. 9. SEPTEMBER 2013

Data Quality: Some Comments on the NASA Software Defect Datasets

Martin Shepperd, Qinbao Song, Zhongbin Sun, and Carolyn Mair

Abstract—Background—Self-evidently empirical analyses rely upon the quality of their data. Likewise, replications rely upon accurate reporting and using the same rather than similar versions of datasets. In recent years, there has been much interest in using machine learners to classify software modules into defect-prone and not defect-prone categories. The publicly available NASA datasets have been extensively used as part of this research. Objective—This short note investigates the extent to which published analyses based on the NASA defect datasets are meaningful and comparable. Method—We analyze the five studies published in the IEEE Transactions on Software Engineering since 2007 that have utilized these datasets and compare the two versions of the datasets currently in use. Results—We find important differences between the two versions of the datasets, implausible values in one dataset and generally insufficient detail documented on dataset preprocessing. Conclusions—It is recommended that researchers 1) indicate the provenance of the datasets they use, 2) report any preprocessing in sufficient detail to enable meaningful replication, and 3) invest effort in understanding the data prior to applying machine learners.

Index Terms—Empirical software engineering, data quality, machine learning, defect prediction

1 Introduction

to classify software modules as faulty or not faulty.

RESENTLY, there is a good deal of interest in using researchers (e.g., Hall et al. [7] found more than a quarter of ■ machine learning methods to induce prediction systems relevant defect prediction studies, that is, 58 out of 208, made use of the NASA datasets). Therefore, these concerns Accurate prediction is useful because it enables among about data integrity and inconsistencies between different

² Shepperd, M., et al., (2013). Data quality: Some comments on the NASA software defect datasets. IEEE TSE, 39(9), 1208-1215.

tl;dr

At least 100 research papers were based on public software defect data sets, until one day ...

we looked more carefully and realised that the data sets made no sense!

COUNT COUN	STEAD HALSTE HALST HALSTEA HALSTE MAINTE MODIFIE	ESSENT ESSEN LOC E PARAM GLOB GLOBAL HALSTE HALSTE HALSTEA HALSTEAD HALSTE HALST HALSTEA HALSTE MAINTE MODIFIE MULTIPL NOI	NORMALI NUM OP NUM NUM
County C			
No.			
4 8 3 0.18 4 2 2 2 6.7 10 2 1 14 0 0 0 0 15.83 13.82 392.38 0.07 6 6 0.07 15.91 21.87 1 2 4 9 0.09 10 10 10 0 4 0 12 0 12 4 2.5 3 0.75 16 3 0.07 10 20 12 10 10 4 0 10 10 10 10 10 10 10 10 10 10 10 10 1			
11 10 4 0.12 4 0.25 3 0.75 16 3 0.07 29 2 0 0 0 9.72 1 0.35 1980, 34 0.26 20 0.75 3 5 14 0.04 50 150 2 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0	0.12 70 0.1 208.38 364.66 0.5 1	1 0 18 1 0 0 35.45 10.29 3750.81 0.12 70 0.1 208.38 364.66 0.5 1 2	0.05 27 43
79 S 98 19 0.17 S 42 2-Q 7 0.39 67 13 0.71 101 2 0 0 83.05 41.39 9999.77 0.8 372 0.02 551.69 341.67 0.72 17 29 53 0.06 150 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.07 46 0.07 167.91 218.72 1 2	3 1 14 0 0 0 15.83 13.82 3022.38 0.07 46 0.07 167.91 218.72 1 2 4	0.09 19 27
Set 40 12 0.17 18 2.22 7 0.58 49 5 0.58 68 5 0.0 0.52.05 2.77 2.9454.09 0.38 1870 0.05 33.5 113.03 0.42 11 20 39 0.07 76 1 1 1 1 1 1 1 1 1	0.26 133 0.05 826.8 768.92 0.75 3	3 0.67 29 2 0 0 39.73 19.35 14882.34 0.26 133 0.05 826.8 768.92 0.75 3 5	4 0.04 50 83
0 1 0 1 0.13 07 1 1 1 2 1 0 7 2 0 0 9.66 2 56.19 0.01 7 0.5 2 16.00 1 0 0 3 0.09 2 1 10 1 0 0 3 0.09 2 1 10 12 4 1 16 1 0 0 71 0 0 0 0 1 17 17 126 110 1 0 10 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 1 1 0 0 0 0 1 0 1 1 0 0 0 1 1 0 0 0 0 1 0 1 0 1 1 0 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 1 0 1 0 1 0 0 0 0 0 1 0 1 0 1 0 0 0 0 1 0 1 0 1 0 0 0 0 1 0 1 0 1 0 0 0 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 0 0 1 0 1 0 1 0 0 1 0 1 0 1 0 0 1 0 1 0 1 0 0 1 0 1 0 1 0 1 0 0 1 0 1 0 1 0 1 0 0 1	0.8 372 0.02 5551.65 2414.97 0.72 17	13 0.71 104 2 0 0 58.36 41.38 99929.77 0.8 372 0.02 5551.65 2414.97 0.72 17 29	1 0.08 150 222
	0.38 187 0.05 1363.54 1130.3 0.42 11	5 0.36 68 5 0 0 52.05 21.71 24543.69 0.38 187 0.05 1363.54 1130.3 0.42 11 20	9 0.07 76 111
22 28 8 0.1 14 2 7 0.88 46 1 0 71 6 0 0 0.9.65 21.96 3771.41 0.66 307 0.65 268.97 1993 0.13 7 14 40 0.06 189 1993 0.13 7 14 40 0.06 189 1993 0.13 7 14 40 0.06 189 1993 0.13 7 14 40 0.06 189 1993 0.13 7 14 40 0.06 189 1993 0.13 7 14 40 0.06 189 1993 0.13 7 14 40 0.06 189 1993 0.13 7 14 40 0.06 189 1993 0.13 7 14 40 0.06 189 1993 0.13 7 14 14 0.10 189 1993 0.14 189 1893 0.14 189 1893 0.14 189 1893 0.14 189 1893 0.14 189 1893 0.14 0.14 0	0.01 7 0.5 2 18.09 1 0	1 0 7 2 0 0 9.05 2 36.19 0.01 7 0.5 2 18.09 1 0 0	3 0.09 2 5
	0.19 104 0.06 563.36 567.78 0.25 3	1 0 21 0 0 0 31.79 17.86 10140.57 0.19 104 0.06 563.36 567.78 0.25 3 6	4 0.05 47 57
4	0.66 307 0.05 2428.97 1993 0.13 7	1 0 71 6 0 0 90.85 21.94 43721.41 0.66 307 0.05 2428.97 1993 0.13 7 14	0.06 108 199
28 30 9 0.13 14 2.14 7 0.78 44 4 0.38 666 5 0 0 76.03 29.08 6.3864.99 1.03 17.07 0.57 2.18 1.05 1 0 0 2 0.17 3 1 1 1 1 1 0 3 0 0 0 0 353 1 2.98 53854.99 1.03 1.04 1.05 1 1 0 0 2 0.17 3 1 1 1 1 1 0 3 0 0 0 0 353 1 2.58 533 1.2 1 0 1 1 1 0 0 2 0 1 1 31 1 1 1 0 1 1 1 1 1 1 0 0 7 1 0 1 1 1 1	0.19 114 0.06 492.91 575.06 0.5 1	1 0 22 1 0 0 37.27 15.43 8872.37 0.19 114 0.06 492.91 575.06 0.5 1 2	7 0.03 54 60
0 1 0 33 07 1 1 1 1 1 1 1 0 3 0 0 0 9 9.83 2 2 93.3 0.01 7 0.5 2.18 19.65 1 0 0 2 2 0.17 3 1 1 1 1 1 1 1 0 0 3 0 0 0 9 9.83 2 2 93.3 0.01 7 0.5 2.18 19.65 1 0 0 0 2 2 0.17 3 1 1 1 1 1 1 1 1 0 0 9 9 9 0 0 0 154.17 22.48 5134.47 0.77 340 0.05 2851.38 230.71 0.14 6 6 12 33 0.05 128 2 2 0 0 0 0 154.57 22.48 5134.47 0.77 340 0.05 2851.38 230.71 0.14 6 6 12 33 0.05 128 2 0 0 0 0 154.57 22.48 5134.47 0.77 340 0.05 2851.38 230.71 0.14 6 6 12 33 0.05 128 2 0 0 0 0 154.57 22.48 5134.47 0.77 340 0.05 2851.38 230.71 0.14 6 6 12 33 0.05 128 2 0 0 0 0 154.57 22.48 5134.47 0.77 340 0.17 340 0.05 2851.38 230.71 0.14 6 6 12 33 0.05 128 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.03 21 0.14 29.28 75.28 0.5 1	1 0 10 0 0 0 10.75 7 526.99 0.03 21 0.14 29.28 75.28 0.5 1 2	6 0.17 7 14
12 12 15 15 15 15 15 15	0.73 342 0.03 3548.06 2203.49 0.44 8	4 0.38 66 5 0 0 76.03 28.98 63864.99 0.73 342 0.03 3548.06 2203.49 0.44 8 15	7 0.07 136 206
22 24 7 0.09 12 2 7 1 1 38 1 1 0 59 9 0 0 10 1475 22.14 51324.87 0.77 349 0.05 2851.31 213.7 0.14 6 12 33 0.05 126 2 1 0 1 0 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 0 1 1 0 0 0 46.76 12325 0.01 1 18 0.05 37.05 586.6 0.3 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.01 7 0.5 2.18 19.65 1 0	1 0 3 0 0 0 9.83 2 39.3 0.01 7 0.5 2.18 19.65 1 0 0	2 0.17 3 4
Second Column Second Colum	0.15 84 0.08 307.3 440.83 1 4	5 1 22 2 0 0 35.13 12.55 5531.32 0.15 84 0.08 307.3 440.83 1 4 6	
	0.77 349 0.05 2851.38 2318.71 0.14 6	1 0 59 9 0 0 104.75 22.14 51324.87 0.77 349 0.05 2851.38 2318.71 0.14 6 12	3 0.05 126 223
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0			
0 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0 0 0 0 0	0.12 74 0.1 198.58 366.61 0.33 2	1 0 19 0 0 0 37.6 9.75 3574.45 0.12 74 0.1 198.58 366.61 0.33 2 4	9 0.06 27 47
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0			2 0.13 2 9
0 0 0 0.11 0 7 1 1 1 2 1 0 9 9 2 0 0 13.7 1 5 304.66 0.02 18 0.2 19.04 68.53 11 0 0 0 3 0.09 4 1 0 0 0 0 0.14 0 7 1 1 1 1 0 0 7 2 0 0 11.07 3 99.66 0.02 18 0.02 18 0.03 55 33.22 11 0 0 0 2 0.08 14 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 1 0	0.01 11 0.29 7.11 36.54 1 0	1 0 8 2 0 10.44 3.5 12 9 0.01 11 0.29 7.11 36.54 1 0 0	2 0.11 3 8
0 0 0 0.11 0 7 1 1 1 2 1 0 9 9 2 0 0 13.7 1 5 304.66 0.02 18 0.2 19.04 68.53 11 0 0 0 3 0.09 4 1 0 0 0 0 0.14 0 7 1 1 1 1 0 0 7 2 0 0 11.07 3 99.66 0.02 18 0.02 18 0.03 55 33.22 11 0 0 0 2 0.08 14 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 1 0	0 6.09 36.54 1 0	3 0 5.09 36.54 1 0 0	2 0.11 4 7
0 0 0 0.11 0 7 1 1 1 2 1 0 9 9 2 0 0 13.7 1 5 304.66 0.02 18 0.2 19.04 68.53 11 0 0 0 3 0.09 4 1 0 0 0 0 0.14 0 7 1 1 1 1 0 0 7 2 0 0 11.07 3 99.66 0.02 18 0.02 18 0.03 55 33.22 11 0 0 0 2 0.08 14 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 1 0	5.09 36.54 1 0	18 3 10 2 d.O. 0 5.09 36.54 1 0 0	2 0.11 4 7
0 0 0 0.11 0 7 1 1 1 2 1 0 9 9 2 0 0 13.7 1 5 304.66 0.02 18 0.2 19.04 68.53 11 0 0 0 3 0.09 4 1 0 0 0 0 0.14 0 7 1 1 1 1 0 0 7 2 0 0 11.07 3 99.66 0.02 18 0.02 18 0.03 55 33.22 11 0 0 0 2 0.08 14 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 1 0	0.09 0. 3.57 261.33 0.5 1	0 19 0	8 0.09 25 32
0 0 0 0.11 0 7 1 1 1 2 1 0 9 9 2 0 0 13.7 1 5 304.66 0.02 18 0.2 19.04 68.53 11 0 0 0 3 0.09 4 1 0 0 0 0 0.14 0 7 1 1 1 1 0 0 7 2 0 0 11.07 3 99.66 0.02 18 0.02 18 0.03 55 33.22 11 0 0 0 2 0.08 14 0 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 1 0	0.03 (19 .27 87.57 0.5 1	21 5.4 47 7 0.03 1 1 1 2	6 0.08 6 17
0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0.05 (33.96 155.42 1 0	1 0 0 0	2 0.06 20 22
1 0 0 4 0 2 2 2 2 1 0 0 27.4 7 1342.72 0.06 45 0.14 74.6 191.82 1 1 0 0 2 0.08 14 62 16 2 0 0 2 14 2 19 1.7 55 89 3 10 0 54.69 47.6 123914.45 0.5 1 2 6 0.11 14 62 16 2 0 0 0 1 0 1.0 14 12 0 0 1 1 0 1 1 1 1 1 1 0 0 1 1 1 1 1	0.02 18 0.2 19.04 68.53 1 0	1 0 9 2 0 0 13.71 5 342.66 0.02 18 0.2 19.04 68.53 1 0 0	
28 34 11 0.14 14 2.43 9 0.8 46 11 1 74 1 0 0 61.92 36.74 83581.43 0.76 354 0.03 4643.41 2274.9 1 10 11 17 37 0.08 124 2 2 16 0.13 10 2 6 1 30 1 0 44 3 3 0 0 0 25.24 22.75 13080.92 0.19 117 0.04 75.26 574.11 0.25 0 0 0 9 0.07 52 16 20 6 0.13 10 2 6 1 30 1 0 44 3 3 0 0 0 46.89 18.91 16764.9 0.3 152 0.05 331.88 886.6 0.17 5 10 26 0.08 52 1 10 1 1 1 1 1 0 8 1 1 0 0 0 22.46 6.19 859.9 0.05 34 0.16 47.77 138.97 1 0 0 0 2 0.06 11 0 1 0 1 1 1 1 1 1 1 0 8 1 1 0 0 0 22.46 6.19 859.9 0.05 34 0.16 47.77 138.97 1 0 0 0 2 0.06 11 0 0 1 0 1 0 1 1 1 1 1 1		1 0 7 0 11.07 3 99.66 0.01 10 0.33 5.54 33.22 1 0 0	
28 34 11 0.14 14 2.43 9 0.8 46 11 1 74 1 0 0 61.92 36.74 83581.43 0.76 354 0.03 4643.41 2274.9 1 10 11 17 37 0.08 124 2 2 16 0.13 10 2 6 1 30 1 0 44 3 3 0 0 0 25.24 22.75 13080.92 0.19 117 0.04 75.26 574.11 0.25 0 0 0 9 0.07 52 16 20 6 0.13 10 2 6 1 30 1 0 44 3 3 0 0 0 46.89 18.91 16764.9 0.3 152 0.05 331.88 886.6 0.17 5 10 26 0.08 52 1 10 1 1 1 1 1 0 8 1 1 0 0 0 22.46 6.19 859.9 0.05 34 0.16 47.77 138.97 1 0 0 0 2 0.06 11 0 1 0 1 1 1 1 1 1 1 0 8 1 1 0 0 0 22.46 6.19 859.9 0.05 34 0.16 47.77 138.97 1 0 0 0 2 0.06 11 0 0 1 0 1 0 1 1 1 1 1 1		0 27.4 7 1342.72 0.06 46 0.14 74.6 191.82 1 0 0	
28 34 11 0.14 14 2.43 9 0.8 46 11 1 74 1 0 0 61.92 36.74 83581.43 0.76 354 0.03 4643.41 2274.9 1 10 11 17 37 0.08 124 2 2 16 0.13 10 2 6 1 30 1 0 44 3 3 0 0 0 25.24 22.75 13080.92 0.19 117 0.04 75.26 574.11 0.25 0 0 0 9 0.07 52 16 20 6 0.13 10 2 6 1 30 1 0 44 3 3 0 0 0 46.89 18.91 16764.9 0.3 152 0.05 331.88 886.6 0.17 5 10 26 0.08 52 1 10 1 1 1 1 1 0 8 1 1 0 0 0 22.46 6.19 859.9 0.05 34 0.16 47.77 138.97 1 0 0 0 2 0.06 11 0 1 0 1 1 1 1 1 1 1 0 8 1 1 0 0 0 22.46 6.19 859.9 0.05 34 0.16 47.77 138.97 1 0 0 0 2 0.06 11 0 0 1 0 1 0 1 1 1 1 1 1		2 0 18.54 9.63 1717.22 0.06 42 0.1 95.4 178.41 0.5 1 2	
28 34 11 0.14 14 2.43 9 0.8 46 11 1 74 1 0 0 61.92 36.74 83591.43 0.76 354 0.03 4643.41 2274.9 1 10 11 73 7 0.08 124 2 2 16 0.13 0 7 4 11 11 1 1 0 0 32 3 0 0 0 25.24 22.75 13080.92 0.19 117 0.04 75.26 5.74.11 0.25 0 0 0 9 0.07 52 16 20 6 0.13 10 2 6 1 30 1 0 44 3 3 0 0 0 46.89 18.91 16764.9 0.3 152 0.05 331.88 886.6 0.17 5 10 26 0.08 52 1 10 1 1 1 1 1 0 8 1 1 0 0 12.46 6.19 889.9 0.05 34 0.16 47.77 139.97 1 0 0 2 0.06 11 0 1 1 1 1 1 1 1 1 1 1 1		89 3 0 0 54.69 47.6 123914.45 0.87 402 0.02 6884.14 2603.24 0.38 6 8	
28 34 11 0.14 14 2.43 9 0.8 46 11 1 74 1 0 0 61.92 36.74 83591.43 0.76 354 0.03 4643.41 2274.9 1 10 11 73 7 0.08 124 2 2 16 0.13 0 7 4 11 11 1 1 0 0 32 3 0 0 0 25.24 22.75 13080.92 0.19 117 0.04 75.26 5.74.11 0.25 0 0 0 9 0.07 52 16 20 6 0.13 10 2 6 1 30 1 0 44 3 3 0 0 0 46.89 18.91 16764.9 0.3 152 0.05 331.88 886.6 0.17 5 10 26 0.08 52 1 10 1 1 1 1 1 0 8 1 1 0 0 12.46 6.19 889.9 0.05 34 0.16 47.77 139.97 1 0 0 2 0.06 11 0 1 1 1 1 1 1 1 1 1 1 1		0 10 3 0 0 16.21 10.56 1809.36 0.06 39 0.09 100.52 171.3 0.5 1 2	
28 34 11 0.14 14 2.43 9 0.8 46 11 1 74 1 0 0 61.92 36.74 83591.43 0.76 354 0.03 4643.41 2274.9 1 10 11 73 7 0.08 124 2 2 16 0.13 0 7 4 11 11 1 1 0 0 32 3 0 0 0 25.24 22.75 13080.92 0.19 117 0.04 75.26 5.74.11 0.25 0 0 0 9 0.07 52 16 20 6 0.13 10 2 6 1 30 1 0 44 3 3 0 0 0 46.89 18.91 16764.9 0.3 152 0.05 331.88 886.6 0.17 5 10 26 0.08 52 1 10 1 1 1 1 1 0 8 1 1 0 0 12.46 6.19 889.9 0.05 34 0.16 47.77 139.97 1 0 0 2 0.06 11 0 1 1 1 1 1 1 1 1 1 1 1		32 1 0 0 34.91 18.57 12037.08 0.22 121 0.05 668.73 648.26 0.25 3 6	
4 0 4 0.13 0? 4 1 11 1 0 32 3 0 0 0 2.524 22.75 13060.92 0.19 117 0.04 725.61 574.11 0.25 0 0 0 9 0.07 52 16 20 6 0.13 10 2 6 6 1 30 1 0 44 3 0 0 0 46.89 18.91 16764.79 0.3 152 0.05 34 0.16 47.77 15 10 26 0.08 52 1 1 0 1 0 1 1 1 1 1 0 8 1 1 0 0 22.46 6.19 859.9 0.05 34 0.16 47.77 138.97 1 0 0 2 0.06 11 0 1 0 1 1 1 1 1 1 1 0 0 8 1 1 0 0 0 22.46 6.19 859.9 0.05 34 0.16 47.77 138.97 1 0 0 2 0.06 11 0 1 0 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 22.86 6.87 2026.77 0.1 62 0.15 112.6 294.8 0.5 1 1 2 6 0.13 20 0 0 1 0.14 0? 2 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 21.8 5.63 669.76 0.04 30 0.18 38.32 122.62 1 0 0 0 2 0.06 11 0 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1			
16			0100 11 100
2 0 1 0.13 0.7 1 0.13 0.7 1 1 1 1 1 0 8 1 1 0 0 0 22.46 6.19 859.9 0.05 34 0.16 47.77 138.97 1 0 0 0 2 0.06 11 0.14 0.7 1 1 0.1 1 1 1 1 0 0 7 1 1 0 0 2 2 0.13 20 0 0 42.88 6.87 2026.77 0.1 62 0.15 112.6 294.8 0.5 1 2 6 0.13 20 0 0 1 0.14 0.7 1 1 1 1 1 1 0 0 7 1 1 0 0 0 21.8 5.63 699.76 0.04 30 0.18 36.32 12.62 1 1 0 0 0 2 0.13 10 0 0 1 1 0.14 0.7 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 51.85 64.92 218524.86 1.12 516 0.02 12140.27 3366.16 0.5 17 31 59 0.07 193 3 0 0 1 1 0.14 0.7 0 7 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 51.85 64.92 218524.86 1.12 516 0.02 12140.27 3366.16 0.5 17 31 59 0.07 193 3 10 0 0 1 1 0.14 0.7 0 7 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 51.85 64.92 18524.86 1.12 516 0.02 12140.27 3366.16 0.5 17 31 59 0.07 193 3 10 0 0 1 1 0.17 0.7 0 7 1 1 1 1 1 1 1 0 0 7 1 0 0 0 145.8 56.9 699.76 0.04 30 0.18 38.32 122.62 1 1 0 0 0 2 0.13 10 0 0 1 1 0.17 0.7 0 7 1 1 1 1 1 1 1 0 0 7 1 0 0 0 145.8 56.9 699.76 0.04 30 0.18 38.32 122.62 1 1 0 0 0 2 0 0.13 10 0 0 0 1 1 0.14 0.7 0 7 1 1 1 1 1 1 1 0 0 7 1 0 0 0 145.8 56.9 699.76 0.04 30 0.18 38.32 122.62 1 1 0 0 0 2 2 0.13 10 0 0 0 1 1 0.17 0.7 0 7 1 1 1 1 1 1 1 0 0 7 1 0 0 0 145.93 66 699.47 0.03 26 0.17 33.86 101.58 1 0 0 0 2 2 0.13 80 10 0 0 1 1 0.08 0 7 1 1 1 1 1 1 1 0 1 0 12 2 0 0 0 0 45.34 4.23 811.53 0.06 46 0.24 45.09 191.82 1 0 0 0 1 1 0.03 20 0 0 4 2 0.18 10 0 0 1 1 0.03 10 0 0 1 1 0.03 20 0 0 1 1 0.14 0 0 1 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 1 0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1 1 0 0 1 1 0 0 1 1 1 0 0 1 1 0 0 1 1 1 0 0 1 1 0 0 1 1 1 0 0 1 1 1 0 0 1 1 0 0 1 1 1 0 0			
0 4 2 0.13 2 2 1 0.5 6 1 0 0 15 2 0 0 0 42.88 6.87 2026.77 0.1 62 0.15 112.6 294.8 0.5 1 2 6 0.13 20 0 0 1 0.14 0 7 1 1 1 1 1 1 1 0 7 1 1 0 0 7 1 1 0 0 0 21.8 5.63 689.76 0.04 30 0.18 38.32 122.62 1 0 0 0 2 0.13 10 0 0 1 1 0.14 0 7 1 1 1 1 1 1 1 1 0 7 1 1 0 0 7 1 1 0 0 0 51.85 64.92 218524.86 1.12 516 0.02 12140.77 3366.16 0.5 17 31 59 0.07 193 3 0 0 1 1 0.14 0 7 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 21.8 5.63 689.76 0.04 30 0.18 38.32 122.62 1 0 0 0 0 2 0.13 10 0 0 1 1 0.14 0 7 1 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 21.8 5.63 689.76 0.04 30 0.18 38.32 122.62 1 1 0 0 0 2 2 0.13 10 0 0 1 1 0.17 0 7 1 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 2 1.8 5.63 689.76 0.04 30 0.18 38.32 122.62 1 1 0 0 0 2 2 0.13 10 0 0 1 1 0.17 0 7 1 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 2 2 0.13 10 0 0 1 1 0.17 0 7 1 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 2 2 0.13 10 0 0 1 1 0.17 0 7 1 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 2 2 0.13 10 0 0 0 1 1 0.17 0 7 1 1 1 1 1 1 1 1 0 0 1 1 0 0 1 2 0 0 0 0			
0			
73 62 18 0.16 28 2.21 10 0.56 75 9 0.47 107 1 0 0 51.85 64.92 218524.86 1.12 516 0.02 12140.73 3366.16 0.5 17 31 59 0.07 193 3 0 0 1 1 0.14 0 ? 1 1 1 1 1 1 1 0 0 7 1 1 0 0 0 51.85 64.92 218524.86 1.12 516 0.02 12140.73 3366.16 0.5 17 31 59 0.07 193 3 0 0 1 1 0.17 0 ? 1 1 1 1 1 1 1 0 0 6 1 1 0 0 0 16.93 6 609.47 0.03 26 0.17 33.86 101.58 1 0 0 0 2 0.13 18 12 0 1 0.08 0 ? 1 1 1 1 1 1 1 0 1 1 0 12 0 0 0 0 45.34 4.23 811.53 0.06 46 0.24 45.09 191.82 1 0 0 0 111 0.03 22 0 0 4 2 0.14 2 2 2 1 1 6 1 1 0 1 1 0 1 2 2 0 0 0 20.76 2.5 129.72 0.02 15 0.4 45.91 151.89 0.5 1 2 6 0.12 6 0 8 3 0.14 4 2 3 3 1 1 0 1 1 0 1 7 2 0 0 0 34.27 2.78 264.4 0.03 25 0.36 14.69 95.18 0.33 2 4 9 0.12 10 0 12 4 0.24 6 2 1 0.25 14 3 0.67 17 2 0 0 0 10.6 10.13 1086.54 0.04 29 0.1 60.36 107.31 0.75 3 6 12 0.22 9 0 4 2 0.14 2 2 2 1 0.5 5 1 0 14 2 2 0 0 0 11.85 9 959.45 0.04 28 0.11 53.3 10.66 0.5 1 2 5 0.11 10 0 4 2 0.14 2 2 2 1 1 0.5 5 1 0 10 1 0 1 0 0 6.92 8 442.81 0.02 16 0.13 24.6 55.35 0.5 1 2 5 0.15 6 15 42 13 0.17 18 2.33 9 0.69 59 7 0.5 70 1 0 0 8.81 26.6 577.6 9 1 0.72 343 0.04 27 343 0.04 32 42.11 0.25 1 2 2 0 0.0 7.44 1 0 0 8.81 26.6 577.6 9 1 0.72 343 0.04 32 42.11 0.25 1 2 2 0 0.0 7.44 1 0 0 0 14.52 5.79 486 0.03 21 0.17 27 84 1 0 0 0 2 0.06 74 4 1 0 0 0 2 0.06 74 4 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1			
0 0 1 0.14 0 ? 1 0.14 0 ? 1 1 1 1 1 1 0 0 7 1 1 0 0 0 21.8 5.63 689.76 0.04 30 0.18 38.32 122.62 1 0 0 0 2 0.13 10 12 0 1 0.17 0 ? 1 1 1 1 1 1 1 1 0 6 6 1 0 0 16.93 6 609.47 0.03 26 0.17 33.86 101.58 1 0 0 0 2 0.13 8 2 12 0 0 1 0.08 0 ? 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1			
0 0 1 0.17 0 7 1 1 1 1 1 1 0 6 1 1 0 0 0 16.93 6 609.47 0.03 26 0.17 33.86 101.58 1 0 0 0 2 0.13 8 12 0 1 0.08 0 7 1 1 1 1 1 1 0 1 0 12 0 0 0 0 45.34 4.23 811.53 0.66 46 0.24 45.09 191.82 1 0 0 0 11 0.03 22 0 4 2 0.14 2 2 2 2 1 1 6 1 0 12 2 0 0 0 34.27 2.78 264.4 0.03 25 0.35 14.69 95.18 0.33 2 4 9 0.12 10 0 12 4 0.24 6 2 1 0.25 14 3 0.67 17 2 0 0 0 10.6 10.13 1086.54 0.04 29 0.1 60.36 107.31 0.75 3 6 12 0.22 9 0 16 6 0.4 6 2.67 1 0.17 18 4 0.6 15 2 0 0 0 7.45 30.8 7065.15 0.08 54 0.03 392.15 129.39 0.67 5 8 14 0.32 22 0 4 2 0.14 2 2 2 2 1 0.5 5 1 0 14 2 0 0 11.85 9 959.45 0.04 28 0.11 53.3 106.61 0.5 1 2 5 0.11 10 0 4 2 0.14 2 2 2 2 1 0.5 5 1 0 14 2 0 0 0 18.85 9 959.45 0.04 28 0.11 53.3 106.61 0.5 1 2 5 0.15 6 15 42 13 0.17 18 2.33 9 0.69 59 7 0.5 70 1 0 0 0 8.81 26.6 5717.691 0.72 343 0.04 3176.49 2149.51 0.54 12 23 48 0.11 133 2 2 10 1 1 1 1 1 0 0 6 3 0 0 14.52 5.79 486 0.03 21 0.17 2 7 84 1 0 0 0 2 0.06 9 18 3 0.16 4 2 2 2 0.67 10 1 1 0 1 0 12 1 0 0 0 28.99 18.12 9514.39 0.18 106 0.06 528.58 525.14 0.33 2 4 9 0.07 44 10 0 CS5702 > Seminar 3: Engineer or Hack 1 0 5 5 1 0 0 9 1 0 0 13.77 13.5 5208.88 0.06 43 0.07 139.38 185.84 0.5 1 2 5 5 0.1 12 5 5 0.1 18			
12			
0 4 2 0.14 2 2 2 2 1 6 1 0 12 2 0 0 0 20.76 2.5 129.72 0.02 15 0.4 7.21 51.89 0.5 1 2 6 0.12 6 0 8 3 0.14 4 2 3 3 1 10 1 0 1 0 17 2 0 0 0 34.27 2.78 264.4 0.03 25 0.36 14.69 95.18 0.33 2 4 9 9 0.12 10 10 12 4 0.24 6 2 1 0.25 14 3 0.67 17 2 0 0 0 10.6 10.13 1086.54 0.04 29 0.1 60.36 107.31 0.75 3 6 12 0.2 9 10 16 6 0.4 6 2.67 1 0.17 18 4 0.6 15 2 0 0 7.45 30.8 706.51 0.08 54 0.03 392.51 229.39 0.67 5 8 14 0.32 22 1 0.5 5 1 0 14 2 0 0 0 11.85 9 959.45 0.04 28 0.11 53.3 106.61 0.5 1 2 5 0.11 10 10 10 1 0 0 6.92 8 442.81 0.02 16 0.13 24.6 55.35 0.5 1 2 2 5 0.15 6 15 0 1 0 10 1 1 0 0 0 80.81 26.6 5717.61 0.72 343 0.04 3176.49 2149.51 0.54 12 2 3 48 0.11 133 2 12 12 4 0.15 6 2 4 1 17 1 0 0 24 2 0 0 0 26.32 16 6737.71 0.14 85 0.06 374.32 421.11 0.25 3 6 15 0.11 30 12 10 10 10 10 10 10 10 10 10 10 10 10 10			
0 8 3 0.14 4 2 3 1 10 1 0 17 2 0 0 0 34.27 2.78 264.4 0.03 25 0.36 14.69 95.18 0.33 2 4 9 0.12 10 0 12 4 0.24 6 2 1 0.25 14 3 0.67 17 2 0 0 0 10.6 10.13 1086.54 0.04 29 0.1 60.36 107.31 0.75 3 6 12 0.22 9 1 1 1 0.5 1 1 0 1 0 1 1 1 0 1 1 1 0 1 1 1 0 1			
0 12 4 0.24 6 2 1 0.25 14 3 0.67 17 2 0 0 0 10.6 10.13 1086.54 0.04 29 0.1 60.36 107.31 0.75 3 6 12 0.22 9 0 16 6 0.4 6 2.67 1 0.17 18 4 0.6 15 2 0 0 7.45 30.8 7065.15 0.08 54 0.03 392.51 229.39 0.67 5 8 14 0.32 22 0 0 4 2 0.14 2 2 1 1 0.5 5 1 0 14 2 0 0 0 11.85 9 959.45 0.04 28 0.11 53.3 106.61 0.5 1 2 5 0.11 10 10 0 4 2 0.2 2 2 1 0.5 5 1 0 10 1 1 0 0 0 6.92 8 44.81 0.02 16 0.13 24.6 55.35 0.5 1 2 2 3 48 0.11 133 2 12 12 4 0.15 6 2 4 1 17 1 0 24 2 0 0 0 26.32 16 6737.71 0.14 85 0.06 374.32 421.11 0.25 3 6 15 0.11 30 0 0 2 0.06 9 19 8 3 0.16 4 2 2 0 0.67 10 1 0 1 0 1 0 0 0 28.99 18.12 9514.39 0.18 106 0.06 528.58 525.14 0.33 2 4 9 0.07 44 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0			
0 16 6 0.4 6 2.67 1 0.17 18 4 0.6 15 2 0 0 7.45 30.8 7065.15 0.08 54 0.03 392.51 229.39 0.67 5 8 14 0.32 22 0 0 4 2 0.14 2 2 1 0.5 5 1 0 14 2 0 0 0 11.85 9 959.45 0.04 28 0.11 53.3 106.61 0.5 1 2 5 0.11 10 0 0 4 2 0.2 2 2 2 1 0.5 5 1 0 10 10 1 0 0 6.92 8 442.81 0.02 16 0.13 24.6 55.35 0.5 1 2 2 3 48 0.11 133 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1			
0			
0 4 2 0.2 2 2 1 0.5 5 1 0 10 1 1 0 0 6.92 8 442.81 0.02 16 0.13 24.6 55.35 0.5 1 2 5 0.15 6 15 42 13 0.17 18 2.33 9 0.69 59 7 0.5 70 1 0 0 80.81 26.6 57176.91 0.72 343 0.04 3176.49 2149.51 0.54 12 23 48 0.11 133 2 12 12 4 0.15 6 2 4 1 17 1 0 24 2 0 0 0 26.32 16 6737.71 0.14 85 0.06 374.32 421.11 0.25 3 6 15 0.11 30 12 1 0 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1			
15			
2 12 4 0.15 6 2 4 1 17 1 0 24 2 0 0 26.32 16 6737.71 0.14 85 0.06 374.32 421.11 0.25 3 6 15 0.11 30 2 0 1 0.11 0 ? 1 1 1 1 1 0 6 3 0 0 14.52 5.79 486 0.03 21 0.17 27 84 1 0 0 0 2 0.06 9 9 8 3 0.16 4 2 2 0.67 10 1 0 1 0 12 1 0 0 28.99 18.12 9514.39 0.18 106 0.06 528.58 525.14 0.33 2 4 9 0.07 44 1 0 0 0 2 0.08 12 1 0 0 9.83 2 39.3 0.01 7 0.5 2.18 19.65 1 0 0 2 0.08 12 3 1 0 0 0 17.66 10.63 1993.75 0.06 45 0.09 110.76 187.65 1 0 0 2 0.05 17 1 1 1 1 0 8 4 0 0 17.66 10.63 1993.75 0.06 43 0.07 139.38 185.84 0.5 1 2 5 0.1 18			
2 0 1 0.11 0 ? 1 1 1 1 1 0 6 3 0 0 14.52 5.79 486 0.03 21 0.17 27 84 1 0 0 2 0.06 9 9 8 3 0.16 4 2 2 0.67 10 1 0 12 1 0 0 28.99 18.12 9514.39 0.18 106 0.06 528.58 525.14 0.33 2 4 9 0.07 44 1 0 CS5702 > Seminar 3: Engineer or Hack 1 0 5 2 0 0 9.83 2 39.3 0.01 7 0.5 2.18 19.65 1 0 0 2 0.08 12 3 2 0 1 0.08 0 ? 1 1 1 1 0 8 4 0 0 17.66 10.63 1993.75 0.06 45 0.09 110.76 187.65 1 0 0 2 0.05 17 2 4 2 0.18 2 2 1 0.5 5 1 0 9 1 0 0 13.77 13.5 2508.88 0.06 43 0.07 139.38 185.84 0.5 1 2 5 0.1 18			
9 8 3 0.16 4 2 2 0.67 10 1 0 12 1 0 0 28.99 18.12 9514.39 0.18 106 0.06 528.58 525.14 0.33 2 4 9 0.07 44 1 0 CS5702 > Seminar 3: Engineer or Hack 1 0 5 2 0 0 9.83 2 39.3 0.01 7 0.5 2.18 19.65 1 0 0 2 0.08 12 3 2 0 1 0.08 0 ? 1 1 1 1 0 8 4 0 0 17.66 10.63 1993.75 0.06 45 0.09 110.76 187.65 1 0 0 2 0.05 17 2 4 2 0.18 2 2 1 0.5 5 1 0 9 1 0 0 13.77 13.5 2508.88 0.06 43 0.07 139.38 185.84 0.5 1 2 5 0.1 18			
1 0 CS5702 > Seminar 3: Engineer or Hack 1 0 5 2 0 0 9.83 2 39.3 0.01 7 0.5 2.18 19.65 1 0 0 2 0.08 12 3 2 0 1 0.08 0 ? 1 1 1 1 1 0 8 4 0 0 17.66 10.63 1993.75 0.06 45 0.09 110.76 187.65 1 0 0 2 0.05 17 2 4 2 0.18 2 2 1 0.5 5 1 0 9 1 0 0 13.77 13.5 2508.88 0.06 43 0.07 139.38 185.84 0.5 1 2 5 0.1 18			
2 0 1 0.08 0 ? 1 1 1 1 1 0 8 4 0 0 17.66 10.63 1993.75 0.06 45 0.09 110.76 187.65 1 0 0 2 0.05 17 2 4 2 0.18 2 2 1 0.5 5 1 0 9 1 0 0 13.77 13.5 2508.88 0.06 43 0.07 139.38 185.84 0.5 1 2 5 0.1 18			
2 4 2 0.18 2 2 1 0.5 5 1 0 9 1 0 0 13.77 13.5 2508.88 0.06 43 0.07 139.38 185.84 0.5 1 2 5 0.1 18		lack 1 0 5 2 0 0 9.83 2 39.3 0.01 7 0.5 2.18 19.65 1 0 0	
2 4 2 0 19 2 2 1 0 5 5 1 0 0 2 0 0 12 64 17 06 2072 02 0 0 5 20 67 222 70 0 5 1 2 5 0 14 24			

Wisdom of crowds

CS5702 > Seminar 3: Engineer or Hack

Data checking in R

The following RMarkdown notebook is based on a simplified version of one of the software defect data sets described above.

(https://raw.githubusercontent.com/mjshepperd/CS5702-Data/master/CS5702_W3_NASAdata.Rmd) [https://raw.githubusercontent.com/mjshepperd/CS5702-Data/master/CS5702W3NASAdata.Rmd]

3. Debugging and Getting help

- 1. find/read the relevant cheatsheet
- 2. perspiration e.g., see this five step approach
- 3. talk it over with a fellow student
- 4. module FAQs on Blackboard
- 5. Stack overflow
- 6. asking a member of the course team

For more suggestions visit the subsection 0.2 "vi) Learn how to get help" in the Modern Data book.

4. Lab Exercises

4.1 User defined functions

The RMarkdown Week 3 Lab Worksheet is available from:

https://raw.githubusercontent.com/mjshepperd/CS5702-Data/master/CS5702W3Lab.Rmd and also on BBL (module content > Week 3)