# Teaching ethics in a new data science course

David Sterratt

8 February 2021

# Context

- "Inf2 - Foundations of Data Science" (Inf2-FDS, or just FDS) is a new 2nd year undergraduate course, taken by Informatics students on all degree programmes

# Context

THE UNIVERSITY of EDINBURGH
informatics

FOUNDATIONS
OF
DATA
SCIENCE

- "Inf2 - Foundations of Data Science" (Inf2-FDS, or just FDS) is a new 2nd year undergraduate course, taken by Informatics students on all degree programmes
- 330 students in 2020/21

# Context

THE UNIVERSITY of EDINBURGH informatics

FOUNDATIONS OF DATA SCIENCE

- "Inf2 - Foundations of Data Science" (Inf2-FDS, or just FDS) is a new 2nd year undergraduate course, taken by Informatics students on all degree programmes
- 330 students in 2020/21
- Delivered by Kobi Gal and David Sterratt

# Context

- "Inf2 - Foundations of Data Science" (Inf2-FDS, or just FDS) is a new 2nd year undergraduate course, taken by Informatics students on all degree programmes
- 330 students in 2020/21
- Delivered by Kobi Gal and David Sterratt
- Two semester, 20 credit course

# Context

- "Inf2 - Foundations of Data Science" (Inf2-FDS, or just FDS) is a new 2nd year undergraduate course, taken by Informatics students on all degree programmes
- 330 students in 2020/21
- Delivered by Kobi Gal and David Sterratt
- Two semester, 20 credit course
- Arises out of the pre-honours curriculum redesign

# Context

- "Inf2 - Foundations of Data Science" (Inf2-FDS, or just FDS) is a new 2nd year undergraduate course, taken by Informatics students on all degree programmes
- 330 students in 2020/21
- Delivered by Kobi Gal and David Sterratt
- Two semester, 20 credit course
- Arises out of the pre-honours curriculum redesign
- Push to embed ethics in the curriculum

# Intended learning outcomes

1. Describe and apply good practices for storing, manipulating, summarising, and visualising data. (**Data storage, manipulation and visualisation**)

# Intended learning outcomes

1. Describe and apply good practices for storing, manipulating, summarising, and visualising data. (**Data storage, manipulation and visualisation**)
2. Use standard packages and tools for data analysis and describing this analysis, such as Python and LaTeX. (**Tool use**)

# Intended learning outcomes

1. Describe and apply good practices for storing, manipulating, summarising, and visualising data. (**Data storage, manipulation and visualisation**)
2. Use standard packages and tools for data analysis and describing this analysis, such as Python and LaTeX. (**Tool use**)
3. Apply basic techniques from descriptive and inferential statistics and machine learning; interpret and describe the output from such analyses. (**Stats and ML** )

# Intended learning outcomes

1. Describe and apply good practices for storing, manipulating, summarising, and visualising data. (**Data storage, manipulation and visualisation**)
2. Use standard packages and tools for data analysis and describing this analysis, such as Python and LaTeX. (**Tool use**)
3. Apply basic techniques from descriptive and inferential statistics and machine learning; interpret and describe the output from such analyses. (**Stats and ML** )
4. *Critically evaluate data-driven methods and claims from case studies, in order to identify and discuss a) potential ethical issues* and b) the extent to which stated conclusions are warranted given evidence provided. (**Critical evaluation**)

# Intended learning outcomes

1. Describe and apply good practices for storing, manipulating, summarising, and visualising data. (**Data storage, manipulation and visualisation**)
2. Use standard packages and tools for data analysis and describing this analysis, such as Python and LaTeX. (**Tool use**)
3. Apply basic techniques from descriptive and inferential statistics and machine learning; interpret and describe the output from such analyses. (**Stats and ML** )
4. *Critically evaluate data-driven methods and claims from case studies, in order to identify and discuss a) potential ethical issues* and b) the extent to which stated conclusions are warranted given evidence provided. (**Critical evaluation**)
5. Complete a data science project and write a report describing the question, methods, and results. (**Data science project**)

# Course description - technical topics

1. Data wrangling and exploratory data analysis

# Course description - technical topics

1. Data wrangling and exploratory data analysis
    - Working with tabular data

# Course description - technical topics

1. Data wrangling and exploratory data analysis
    - Working with tabular data
    - Descriptive statistics and visualisation

# Course description - technical topics

1. Data wrangling and exploratory data analysis
    - Working with tabular data
    - Descriptive statistics and visualisation
    - Linear regression and correlation

# Course description - technical topics

1. Data wrangling and exploratory data analysis
   - Working with tabular data
   - Descriptive statistics and visualisation
   - Linear regression and correlation
   - Clustering

# Course description - technical topics

1. Data wrangling and exploratory data analysis
   - Working with tabular data
   - Descriptive statistics and visualisation
   - Linear regression and correlation
   - Clustering
2. Supervised machine learning

# Course description - technical topics

1. Data wrangling and exploratory data analysis
    - Working with tabular data
    - Descriptive statistics and visualisation
    - Linear regression and correlation
    - Clustering
2. Supervised machine learning
    - Classification

# Course description - technical topics

1. Data wrangling and exploratory data analysis
   - Working with tabular data
   - Descriptive statistics and visualisation
   - Linear regression and correlation
   - Clustering
2. Supervised machine learning
   - Classification
   - More on linear regression; logistic regression

# Course description - technical topics

1. Data wrangling and exploratory data analysis
   - Working with tabular data
   - Descriptive statistics and visualisation
   - Linear regression and correlation
   - Clustering
2. Supervised machine learning
   - Classification
   - More on linear regression; logistic regression
   - Generalization and regularization

# Course description - technical topics

1. Data wrangling and exploratory data analysis
   - Working with tabular data
   - Descriptive statistics and visualisation
   - Linear regression and correlation
   - Clustering
2. Supervised machine learning
   - Classification
   - More on linear regression; logistic regression
   - Generalization and regularization
3. Statistical inference

# Course description - technical topics

1. Data wrangling and exploratory data analysis
   - Working with tabular data
   - Descriptive statistics and visualisation
   - Linear regression and correlation
   - Clustering
2. Supervised machine learning
   - Classification
   - More on linear regression; logistic regression
   - Generalization and regularization
3. Statistical inference
   - Randomness, simulation and sampling

# Course description - technical topics

1. Data wrangling and exploratory data analysis
    - Working with tabular data
    - Descriptive statistics and visualisation
    - Linear regression and correlation
    - Clustering
2. Supervised machine learning
    - Classification
    - More on linear regression; logistic regression
    - Generalization and regularization
3. Statistical inference
    - Randomness, simulation and sampling
    - Confidence intervals, law of large numbers

# Course description - technical topics

1. Data wrangling and exploratory data analysis
   - Working with tabular data
   - Descriptive statistics and visualisation
   - Linear regression and correlation
   - Clustering
2. Supervised machine learning
   - Classification
   - More on linear regression; logistic regression
   - Generalization and regularization
3. Statistical inference
   - Randomness, simulation and sampling
   - Confidence intervals, law of large numbers
   - Randomized studies, hypothesis testing

# Course description - real-world implications

A. Implications:

- *Where does data come from? (Sample bias, data licensing and privacy issues)*

B. Thinking, working, and writing:

# Course description - real-world implications

A. Implications:

- *Where does data come from? (Sample bias, data licensing and privacy issues)*
- *Visualisation: misleading plots, accessible design*

B. Thinking, working, and writing:

# Course description - real-world implications

A. Implications:

- *Where does data come from? (Sample bias, data licensing and privacy issues)*
- *Visualisation: misleading plots, accessible design*
- *Machine learning: algorithmic bias and discrimination*

B. Thinking, working, and writing:

# Course description - real-world implications

A. Implications:

- *Where does data come from? (Sample bias, data licensing and privacy issues)*
- *Visualisation: misleading plots, accessible design*
- *Machine learning: algorithmic bias and discrimination*

B. Thinking, working, and writing:

- Claims and evidence: what can we conclude; analysis of errors

# Course description - real-world implications

A. Implications:

- *Where does data come from? (Sample bias, data licensing and privacy issues)*
- *Visualisation: misleading plots, accessible design*
- *Machine learning: algorithmic bias and discrimination*

B. Thinking, working, and writing:

- Claims and evidence: what can we conclude; analysis of errors
- Reproducibility; programming “notebooks” vs modular code

# Course description - real-world implications

A. Implications:

- *Where does data come from? (Sample bias, data licensing and privacy issues)*
- *Visualisation: misleading plots, accessible design*
- *Machine learning: algorithmic bias and discrimination*

B. Thinking, working, and writing:

- Claims and evidence: what can we conclude; analysis of errors
- Reproducibility; programming “notebooks” vs modular code
- Scientific communication; structure of a lab report

# Course description - real-world implications

A. Implications:

- *Where does data come from? (Sample bias, data licensing and privacy issues)*
- *Visualisation: misleading plots, accessible design*
- *Machine learning: algorithmic bias and discrimination*

B. Thinking, working, and writing:

- Claims and evidence: what can we conclude; analysis of errors
- Reproducibility; programming "notebooks" vs modular code
- Scientific communication; structure of a lab report
- Reading and critique of data science articles

# How it started...

## . . . how it's going

# Explicit ethics content

- S1 week 3: Group presentation and discussion

# Explicit ethics content

- S1 week 3: Group presentation and discussion
- Lectures in S1 week 3: Re-identification, Rights under GDPR, Algorithmic bias (COMPAS, twitter chatbot)

# Explicit ethics content

- S1 week 3: Group presentation and discussion
- Lectures in S1 week 3: Re-identification, Rights under GDPR, Algorithmic bias (COMPAS, twitter chatbot)
- S1: Misleading plots when discussing visualisation

# Explicit ethics content

- S1 week 3: Group presentation and discussion
- Lectures in S1 week 3: Re-identification, Rights under GDPR, Algorithmic bias (COMPAS, twitter chatbot)
- S1: Misleading plots when discussing visualisation
- S2: Critical evaluation exercise had question on ethics of case study, and involved comparison with a media article

# Explicit ethics content

- S1 week 3: Group presentation and discussion
- Lectures in S1 week 3: Re-identification, Rights under GDPR, Algorithmic bias (COMPAS, twitter chatbot)
- S1: Misleading plots when discussing visualisation
- S2: Critical evaluation exercise had question on ethics of case study, and involved comparison with a media article
- S2: When comparing supervised ML methods, discussion of transparency of methods

# Explicit ethics content

- S1 week 3: Group presentation and discussion
- Lectures in S1 week 3: Re-identification, Rights under GDPR, Algorithmic bias (COMPAS, twitter chatbot)
- S1: Misleading plots when discussing visualisation
- S2: Critical evaluation exercise had question on ethics of case study, and involved comparison with a media article
- S2: When comparing supervised ML methods, discussion of transparency of methods
- S2 week 4: Invited lecture by Dr Galina Andreeva from the Business School on interactions between law and outcomes in credit approval

# Explicit ethics content

- S1 week 3: Group presentation and discussion
- Lectures in S1 week 3: Re-identification, Rights under GDPR, Algorithmic bias (COMPAS, twitter chatbot)
- S1: Misleading plots when discussing visualisation
- S2: Critical evaluation exercise had question on ethics of case study, and involved comparison with a media article
- S2: When comparing supervised ML methods, discussion of transparency of methods
- S2 week 4: Invited lecture by Dr Galina Andreeva from the Business School on interactions between law and outcomes in credit approval
- Questions on ethics and law in class tests

# Explicit ethics content

- S1 week 3: Group presentation and discussion
- Lectures in S1 week 3: Re-identification, Rights under GDPR, Algorithmic bias (COMPAS, twitter chatbot)
- S1: Misleading plots when discussing visualisation
- S2: Critical evaluation exercise had question on ethics of case study, and involved comparison with a media article
- S2: When comparing supervised ML methods, discussion of transparency of methods
- S2 week 4: Invited lecture by Dr Galina Andreeva from the Business School on interactions between law and outcomes in credit approval
- Questions on ethics and law in class tests
- S2 week 6 lab: Legal/ethical web scraping

# Ethics discussions in workshop

- Before time, split workshop groups of 16 into 4 “collaboration groups” of 4

# Ethics discussions in workshop

- Before time, split workshop groups of 16 into 4 “collaboration groups” of 4
- Effort put in to gender balance of collaboration groups

# Ethics discussions in workshop

- Before time, split workshop groups of 16 into 4 “collaboration groups” of 4
- Effort put in to gender balance of collaboration groups
- Set task of:

# Ethics discussions in workshop

- Before time, split workshop groups of 16 into 4 "collaboration groups" of 4
- Effort put in to gender balance of collaboration groups
- Set task of:
  1. Read Parts 1 and 2 of Shannon Vallor's *An Introduction to Data Ethics* (on harms and benefits of data, and common ethical challenges of data)

# Ethics discussions in workshop

- Before time, split workshop groups of 16 into 4 “collaboration groups” of 4
- Effort put in to gender balance of collaboration groups
- Set task of:
    1. Read Parts 1 and 2 of Shannon Vallor’s *An Introduction to Data Ethics* (on harms and benefits of data, and common ethical challenges of data)
    2. Read short case study on either

# Ethics discussions in workshop

- Before time, split workshop groups of 16 into 4 “collaboration groups” of 4
- Effort put in to gender balance of collaboration groups
- Set task of:
  1. Read Parts 1 and 2 of Shannon Vallor’s *An Introduction to Data Ethics* (on harms and benefits of data, and common ethical challenges of data)
  2. Read short case study on either
     - Facebook’s emotion manipulation of feed experiment (informed consent, specific harms to users)

# Ethics discussions in workshop

- Before time, split workshop groups of 16 into 4 "collaboration groups" of 4
- Effort put in to gender balance of collaboration groups
- Set task of:
  1. Read Parts 1 and 2 of Shannon Vallor's *An Introduction to Data Ethics* (on harms and benefits of data, and common ethical challenges of data)
  2. Read short case study on either
     - Facebook's emotion manipulation of feed experiment (informed consent, specific harms to users)
     - OK Cupid data breach (legal, terms and conditions)

# Ethics discussions in workshop

- Before time, split workshop groups of 16 into 4 “collaboration groups” of 4
- Effort put in to gender balance of collaboration groups
- Set task of:
  1. Read Parts 1 and 2 of Shannon Vallor’s *An Introduction to Data Ethics* (on harms and benefits of data, and common ethical challenges of data)
  2. Read short case study on either
     - Facebook’s emotion manipulation of feed experiment (informed consent, specific harms to users)
     - OK Cupid data breach (legal, terms and conditions)
  3. Structure a presentation around 6 questions; share presentation with tutor in advance

# Ethics discussions in workshop

- Before time, split workshop groups of 16 into 4 "collaboration groups" of 4
- Effort put in to gender balance of collaboration groups
- Set task of:
  1. Read Parts 1 and 2 of Shannon Vallor's *An Introduction to Data Ethics* (on harms and benefits of data, and common ethical challenges of data)
  2. Read short case study on either
     - Facebook's emotion manipulation of feed experiment (informed consent, specific harms to users)
     - OK Cupid data breach (legal, terms and conditions)
  3. Structure a presentation around 6 questions; share presentation with tutor in advance
  4. Each group presents at 1-hour workshop session in Collaborate, with short time for discussion after each presentation.

# Workshop evaluation

- Good attendance and engagement

# Workshop evaluation

- Good attendance and engagement
- Tutors felt sessions had gone well, though felt that OK Cupid was a rather blatant example

# Workshop evaluation

- Good attendance and engagement
- Tutors felt sessions had gone well, though felt that OK Cupid was a rather blatant example
- However, format did not lead to much discussion in the workshop in some groups - though there was probably discussion beforehand between students

# Student response - positive

- *Data Ethics reading is awesome! I suppose some here may be uninterested in reading about ethics and would way rather focus on the technical parts of Data Science, but after reading the first chapter, I wanted to let you guys know that the reading is very interesting and relevant. As someone with a deep interest in deep learning, a lot of these ethical issues are directly applicable and solving them often involves rather interesting technical solutions, e.g. [. . . ] Even if you do not find ethics interesting, you surely want to understand your models/data, and considering ethics is a part of this process."* (Piazza, at the time of the ethics workshop)

# Student response - positive

- *Data Ethics reading is awesome! I suppose some here may be uninterested in reading about ethics and would way rather focus on the technical parts of Data Science, but after reading the first chapter, I wanted to let you guys know that the reading is very interesting and relevant. As someone with a deep interest in deep learning, a lot of these ethical issues are directly applicable and solving them often involves rather interesting technical solutions, e.g. [. . . ] Even if you do not find ethics interesting, you surely want to understand your models/data, and considering ethics is a part of this process."* (Piazza, at the time of the ethics workshop)
- *I think it is interesting to engage in critical thinking, and sometimes combine it with code.* (S1 mid-semester survey)

# Student response - positive

- *Data Ethics reading is awesome! I suppose some here may be uninterested in reading about ethics and would way rather focus on the technical parts of Data Science, but after reading the first chapter, I wanted to let you guys know that the reading is very interesting and relevant. As someone with a deep interest in deep learning, a lot of these ethical issues are directly applicable and solving them often involves rather interesting technical solutions, e.g. [...] Even if you do not find ethics interesting, you surely want to understand your models/data, and considering ethics is a part of this process."* (Piazza, at the time of the ethics workshop)
- *I think it is interesting to engage in critical thinking, and sometimes combine it with code.* (S1 mid-semester survey)
- *"I wish there was more about ethics"* (Course evaluation questionnaire)

# Extremely negative student response (on Piazza)

*I object to the political bias of this course*
*I thought I'd signed up for a class on Data Science, not Gender Studies or some other left-wing nonsense.*
*Just imagine if the data sets you'd chosen were just as heavily promoting pro-life arguments or insistence on religious education, the dangers of illegal immigration or some other conservative talking point? [...]*
*I did a science degree to get away from the politics constantly being rammed down my throat on a daily basis, and I feel pretty pissed off that you're not only taking every opportunity you can to try to indoctrinate me into this Marxist cult [...]*
*You should remove all politically motivated content from this course as a matter of priority. The University of Edinburgh cannot make any true claim to "diversity" until it removes this political bias from courses that should be apolitical.*

# Next year

- Overall, it felt like a reasonable start but next year...

# Next year

- Overall, it felt like a reasonable start but next year...
- Improve the set-up of discussion in the data ethics workshop

# Next year

- Overall, it felt like a reasonable start but next year...
- Improve the set-up of discussion in the data ethics workshop
- Perhaps one more workshop on ethical issues?

# Next year

- Overall, it felt like a reasonable start but next year...
- Improve the set-up of discussion in the data ethics workshop
- Perhaps one more workshop on ethical issues?
- More discussions in Q&A - perhaps asking students to apply ethical principles to news stories

# Next year

- Overall, it felt like a reasonable start but next year...
- Improve the set-up of discussion in the data ethics workshop
- Perhaps one more workshop on ethical issues?
- More discussions in Q&A - perhaps asking students to apply ethical principles to news stories
- Some student feedback has suggested "harder" topics should be introduced earlier; try to stick with ethics early

# Acknowledgements

- Kobi Gal

# Acknowledgements


- Kobi Gal
- Sharon Goldwater, Heather Yorston

# Acknowledgements


- Kobi Gal
- Sharon Goldwater, Heather Yorston
- Chris Williams

# Acknowledgements

- Kobi Gal
- Sharon Goldwater, Heather Yorston
- Chris Williams
- The marvellous team of FDS tutors and lab demonstrators, including Jonathan Feldstein on Lab development

# Acknowledgements


- Kobi Gal
- Sharon Goldwater, Heather Yorston
- Chris Williams
- The marvellous team of FDS tutors and lab demonstrators, including Jonathan Feldstein on Lab development
- ILTS

# Acknowledgements


- Kobi Gal
- Sharon Goldwater, Heather Yorston
- Chris Williams
- The marvellous team of FDS tutors and lab demonstrators, including Jonathan Feldstein on Lab development
- ILTS
- Other colleagues

# Acknowledgements


- Kobi Gal
- Sharon Goldwater, Heather Yorston
- Chris Williams
- The marvellous team of FDS tutors and lab demonstrators, including Jonathan Feldstein on Lab development
- ILTS
- Other colleagues
- My family