

When to Trust and When Not to Trust Data Science

Saghir Bashir

{i} ilustat
www.ilustat.com

1

Definition: “Data Science”

Generally accepted definition *does not* exist!

Presentation definition:

“Using data, statistics and programming, in a given context, to support decision making.”

“Applied Statistics”

2

Machine Learning is NOT Data Science!

Machine Learning is type of analysis that you might perform as part of doing Data Science

3

Outline

Data Data Everywhere
News Headlines & Data Science
Trust & Trustworthy
Trustworthy Data Science
Summary

{i} ilustat

Objectives

My objectives are to encourage you to:

- > Challenge your own thinking
- > Be objective & critical about Data Science
 - “Trustworthy Data Science”

5

Data Data Everywhere

News Headlines & Data Science

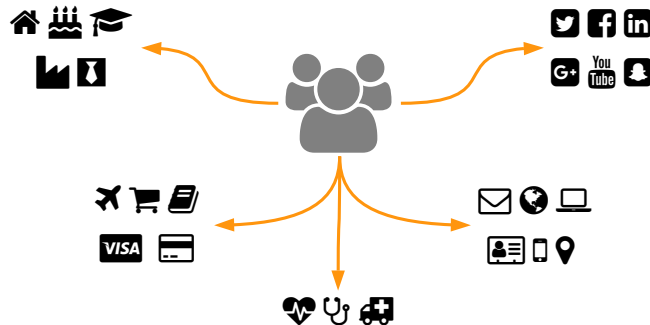
Trust & Trustworthy

Trustworthy Data Science

Summary

{i} ilustrat

We Are Data



7

Data Data Everywhere

> Governmental

→ Unemployment, crime, literacy, economic, census, demographic, ...

> Non-governmental

→ Homelessness, social inequality, poverty, opinion polls and surveys, ...

> Business

→ Stock price, sales data, profits, business confidence, ...

> Internet

→ Social media, search history, web browsing, surveys, ...

> Health

→ Disease monitoring, pharmaceutical, live births, ...

> Environmental

→ Climate, marine, animal, plants, pollution, ...

> And so on...

8

Data Science Everywhere

If we have **DATA** everywhere then
we have **DATA SCIENCE** everywhere

9

Data Data Everywhere

News Headlines & Data Science

Trust & Trustworthy

Trustworthy Data Science

Summary

{i} ilustrat

MailOnline

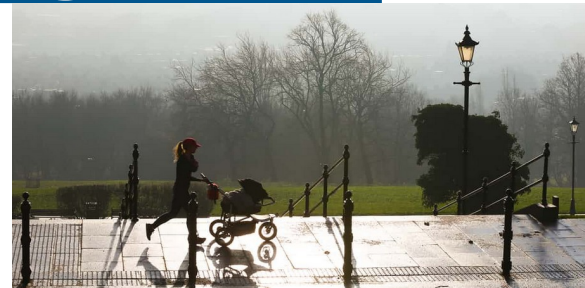
Lethal legacy of dash for diesel: Air pollution is 'killing 40,000 a year in the UK'

- Diesel cars fuel a health crisis that kills 40,000 people a year in the UK
- Emissions linked to asthma, heart disease, cancer, diabetes and dementia
- Ownership of diesel cars has more than trebled in the past 15 years

23 February 2016
<https://1n.pm/gTkce>

11

theguardian



Air pollution crisis 'plagues' UK, finds UN human rights expert

'Silent pandemic' of air pollution affects UK children and there is no indication protection against toxic waste will be retained after Brexit

12

"...between 30,000 and 40,000 early deaths every year are caused by toxic air across the country"

31 January 2017
<https://1n.pm/0Bkqg>

13

AMBIENTE

Poluição do ar provoca 6630 mortes prematuras em Portugal

Lisboa, Porto e Braga encontram-se entre as cidades com valores anuais de poluentes atmosféricos nocivos acima da média, alertam os ambientalistas

MARIA WILTON · 11 de Outubro de 2017, 18:49

508 PARTILHAS [f](#) [t](#) [in](#) [G+](#)



11 October 2017
<https://1n.pm/Yvf8y>

14

Switzerland is 'world's happiest' country in new poll

© 24 April 2015 | Business | [f](#) [t](#) [b](#) [e](#) [Share](#)



24 April 2015
<https://1n.pm/Lglq2>

15

Denmark the 'happiest country' and Burundi 'the least happy'

© 16 March 2016 | Europe | [f](#) [t](#) [b](#) [e](#) [Share](#)



16 March 2016
<https://1n.pm/TBfxE>

16

Happiness report: Norway is the happiest place on Earth

20 March 2017 | World

Share



20 March 2017
<https://1n.pm/EloSY>

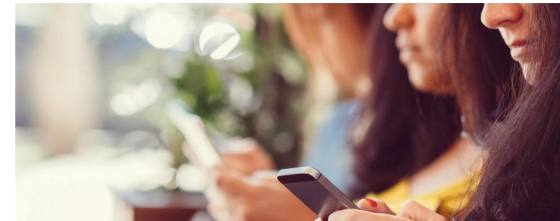
17

News > World > Americas

Users fear social media is making them ill, but they still can't stop

Some 90 per cent of 18 to 29-year-olds now own a smartphone

Rebecca Flood | Sunday 26 February 2017 17:56 GMT



26 February 2017
<https://1n.pm/jPt09>

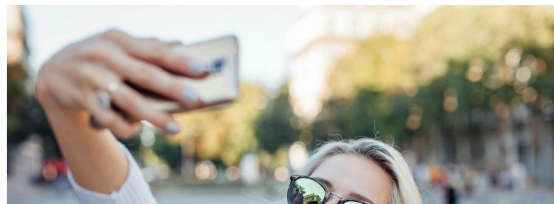
18

Voices

Don't listen to the moral panic, social media is good for young people's mental health

Social media can allow young people to express themselves and build communities. Policing it will only make teenagers withdraw further

Jay Watts | Saturday 20 May 2017 17:52 BST



20 May 2017
<https://1n.pm/EC50i>

19

Some Comments

News headlines

- > Catch your attention and give you a flavour of the story
- > The news story may or may not represent the source
- > People will have different views and interpretations
 - This is not of interest in this presentation
 - The interest is in the "validity and quality" of the source Data Science

20

It's Not About Headlines

One day it could be your Data Science 🤖

- > Perhaps not as a news story with a creative headline
- > Perhaps as a summary to your bosses or clients
- > Could you defend your work?

Can we “trust” the source Data Science?

- > If not, outcomes could be harmful

21

Data Science & Trust

“Air pollution causes 6630 premature deaths in Portugal”

- > What would make you “trust” this headline?

Think about:

- > Bias prevention & reduction measures
- > Characterisation of uncertainty
- > Validity and quality

22

Data Data Everywhere
News Headlines & Data Science
Trust & Trustworthy
Trustworthy Data Science
Summary

{i} ilustrat

Trust

Some thoughts...

- > Earning trust is hard but it is very easy to lose
- > It is not binary
 - Could trust data but not the analysis
- > Data Science has many potential points of “trust failures” and “trust leaks”

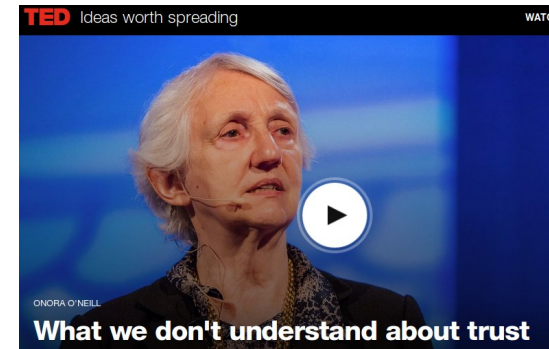
24

Trust & Trustworthy

- > It is not about “trust” or “building trust”
 - Con artists and fraudsters use “trust” to cheat you
 - “Building trust” or “increasing trust” is their art
- > It is about being “trustworthy”
 - Competent
 - Reliable
 - Honest
- > You must EARN trust to be trustworthy
 - You should not just expect to receive it

25

Must Watch! (10 mins)



https://www.ted.com/talks/onora_o_neill_what_we_don_t_understand_about_trust

26

Data Data Everywhere
News Headlines & Data Science
Trust & Trustworthy
Trustworthy Data Science
Summary

{i} ilustat

Open & Transparent

- “Trustworthy Data Science” includes being:**
- > **Open and Transparent**
 - > **Honest** especially about strengths AND weaknesses!
 - > **Willing to do the same as what you expect of others**
 - You cannot set higher standards for others compared to yourself

28

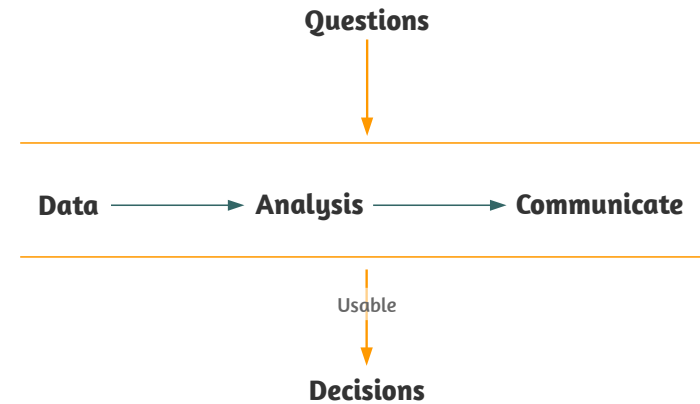
Trustworthy Data Science?

The following slides give some ideas on how to achieve or assess trustworthiness

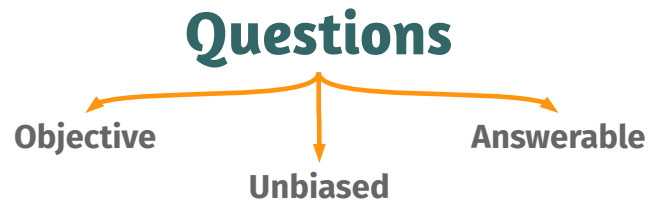
- > They are not a comprehensive list and they are not intended to be
- > “It’s a little bit more complicated than that!” 😊

29

Data Science In Practice

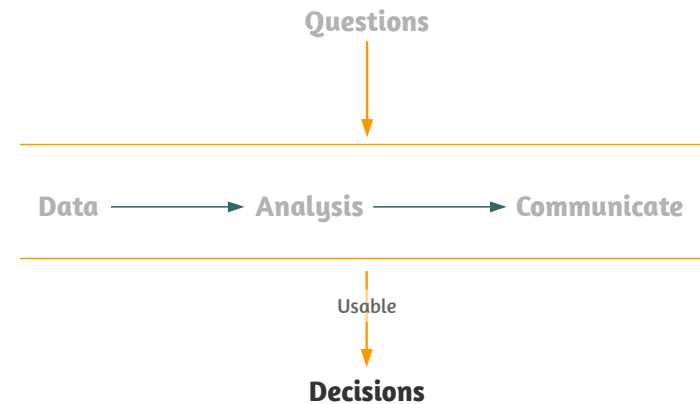


30

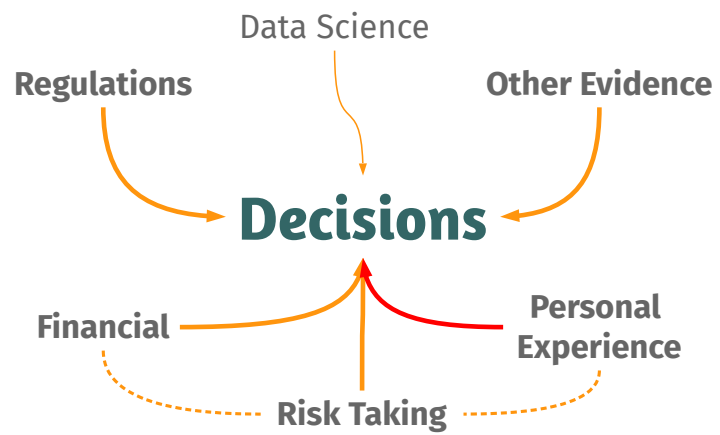


“Are there more deaths from air pollution ?”
vs
“What are the air pollution trends for Portugal?”
“What are the health effects of air pollution?”

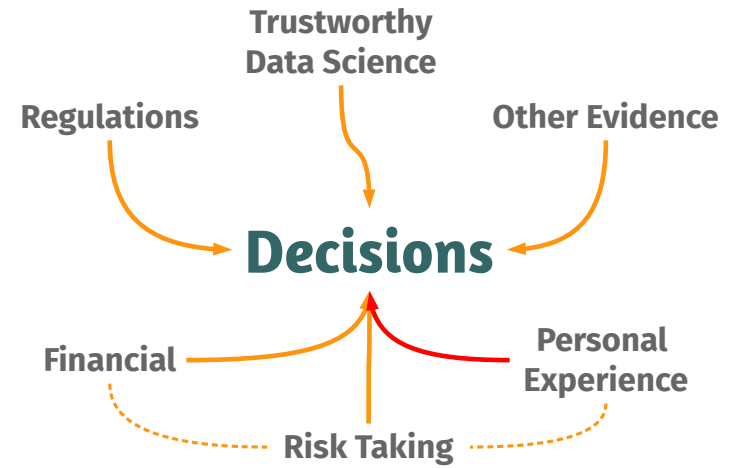
31



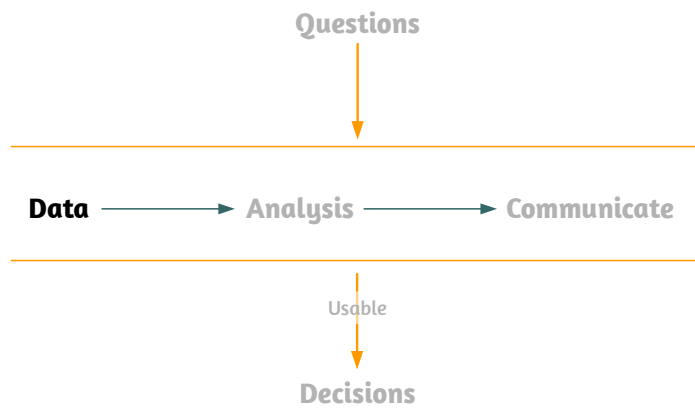
32



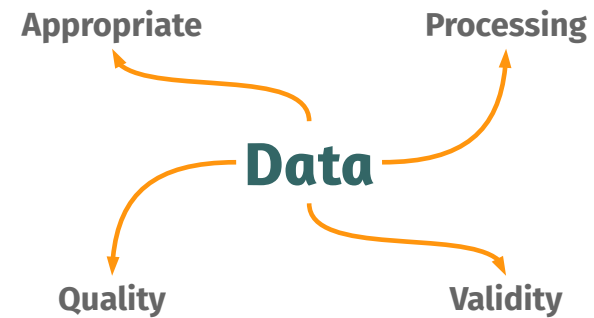
33



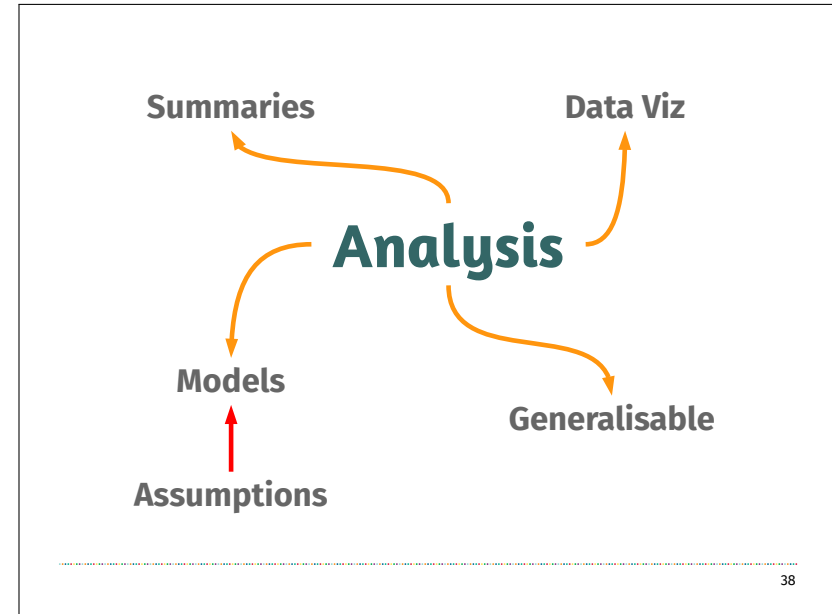
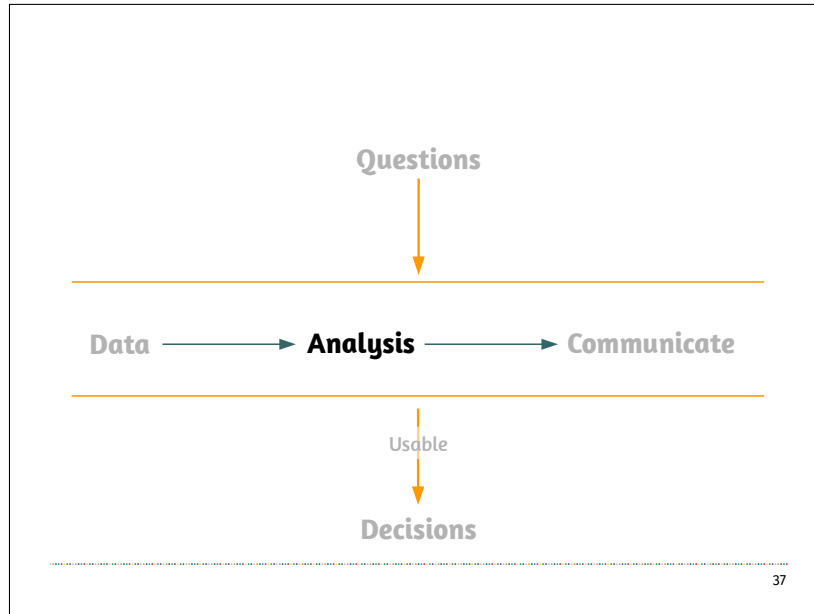
34



35



36

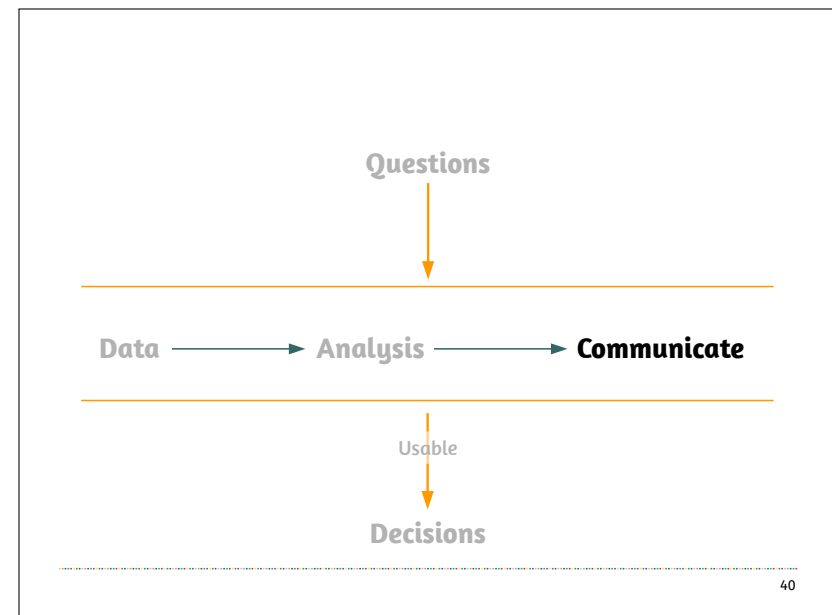


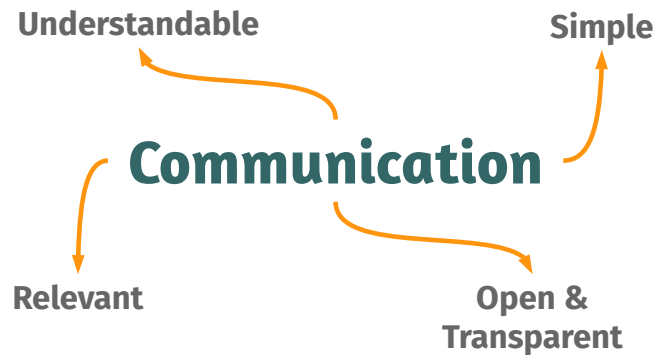
Quotes

"Remember that all models are wrong; the practical question is how wrong do they have to be to not be useful."
George E.P. Box (1987)

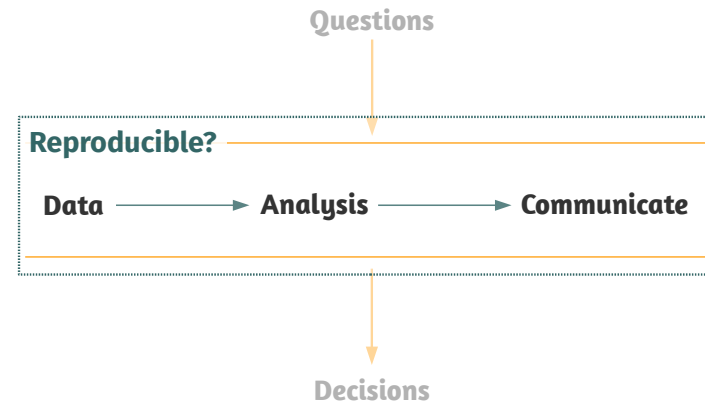
"Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise."
John W. Tukey (1962)

39





41



42

Data Data Everywhere
 News Headlines & Data Science
 Trust & Trustworthy
 Trustworthy Data Science
Summary

{i} ilustat

Summary

When to Trust and When Not to Trust Data Science?

> It is about “trustworthiness”

→ Competent, Reliable & Honest

> Trust must be EARNED to be trustworthy

→ Bias prevention & reduction measures, state uncertainties, ...

→ Openness & transparency about strengths and weaknesses

> “Trustworthy Data Science”

→ Objective and critical evaluation of your work and that of others

→ Reproducibility is an important part but it is not the whole

44

Thank you

Saghir Bashir

 ilustat
www.ilustat.com

45

References

- > ***“What we don’t understand about trust”, Onora O’Neill***
 - > https://www.ted.com/talks/onora_o_neill_what_we_don_t_understand_about_trust
 - > Short link: <https://1n.pm/PhP7>
- > Box, George E. P. & Norman R. Draper (1987). “Empirical Model-Building and Response Surfaces”, Wiley.
- > John W. Tukey (1962). “The future of data analysis”, Annals of Mathematical Statistics 33: 1-67

When to Trust and not to Trust Data Science

46

This work is licensed under the
**Creative Commons Attribution-NonCommercial 4.0
International License.**

To view a copy of this license, visit
<http://creativecommons.org/licenses/by-nc/4.0/>

