

Data scientists supporting scientific researchers: A win-win situation

Matteo Tanadini
Zurich Data Scientists

30th March 2022



Outline

- 1 Who we are
- 2 Aim of the talk
- 3 Use case
- 4 Conclusions

Section 1

Who we are

- Zurich Data Scientists

- Zurich Data Scientists
- Team

- Zurich Data Scientists
- Team
- Matteo Tanadini

- Zurich Data Scientists
- Team
- Matteo Tanadini
- Clients

- Zurich Data Scientists
- Team
- Matteo Tanadini
- Clients
 - ▶ universities (teaching)

- Zurich Data Scientists
- Team
- Matteo Tanadini
- Clients
 - ▶ universities (teaching)
 - ▶ research groups

- Zurich Data Scientists
- Team
- Matteo Tanadini
- Clients
 - ▶ universities (teaching)
 - ▶ research groups
 - ▶ private companies and governmental agencies

Section 2

Aim of the talk

Aim of the talk

- elaborate on benefits and challenges of the data scientists-researchers collaboration

Aim of the talk

- elaborate on benefits and challenges of the data scientists-researchers collaboration
- via a use case

Aim of the talk

- elaborate on benefits and challenges of the data scientists-researchers collaboration
- via a use case
- elaborate on **our role** and how it should evolve

Section 3

Use case

Introducing the use case

Aim: Estimate the risk of local transmission via tiger mosquitos of airborne viruses



Introducing the use case

Aim: Estimate the risk of local transmission via tiger mosquitos of airborne viruses



- tiger mosquito

Introducing the use case

Aim: Estimate the risk of local transmission via tiger mosquitos of airborne viruses



- tiger mosquito
- can help exotic viruses spread locally (eg. chikungunya and Dengue viruses)

Introducing the use case

Aim: Estimate the risk of local transmission via tiger mosquitos of airborne viruses



- tiger mosquito
- can help exotic viruses spread locally (eg. chikungunya and Dengue viruses)
- human intervention is required in high risk situations

“Gold standard” approach

“Gold standard” approach

- quantify abundance of tiger mosquito adults via *human landing collections*

“Gold standard” approach

- quantify abundance of tiger mosquito adults via *human landing collections*
- introduce observed adults abundances in R_0 formulas

“Gold standard” approach

- quantify abundance of tiger mosquito adults via *human landing collections*
- introduce observed adults abundances in R_0 formulas
- quantify local risk

“Newer” approach



“Newer” approach



- indirectly abundance of tiger mosquito adults via eggs abundances

“Newer” approach



- indirectly abundance of tiger mosquito adults via eggs abundances
- introduce estimated abundances in R0 formulas

“Newer” approach



- indirectly abundance of tiger mosquito adults via eggs abundances
- introduce estimated abundances in R_0 formulas
- quantify local risk

The two approaches to estimate adult abundances

- 1 “gold standard” approach

The two approaches to estimate adult abundances

- 1 “gold standard” approach
 - ▶ based on human landing collections

The two approaches to estimate adult abundances

- 1 “gold standard” approach
 - ▶ based on human landing collections
 - ▶ directly quantifies adult abundances

The two approaches to estimate adult abundances

- 1 “gold standard” approach
 - ▶ based on human landing collections
 - ▶ directly quantifies adult abundances
 - ▶ but expensive and requires an ad-hoc effort

The two approaches to estimate adult abundances

- ① “gold standard” approach
 - ▶ based on human landing collections
 - ▶ directly quantifies adult abundances
 - ▶ but expensive and requires an ad-hoc effort
 - ▶ but unethical in high risk settings

The two approaches to estimate adult abundances

- 1 “gold standard” approach
 - ▶ based on human landing collections
 - ▶ directly quantifies adult abundances
 - ▶ but expensive and requires an ad-hoc effort
 - ▶ but unethical in high risk settings
- 2 “newer” approach

The two approaches to estimate adult abundances

- 1 “gold standard” approach
 - ▶ based on human landing collections
 - ▶ directly quantifies adult abundances
 - ▶ but expensive and requires an ad-hoc effort
 - ▶ but unethical in high risk settings
- 2 “newer” approach
 - ▶ **indirectly** quantifies adult abundances

The two approaches to estimate adult abundances

- 1 “gold standard” approach
 - ▶ based on human landing collections
 - ▶ directly quantifies adult abundances
 - ▶ but expensive and requires an ad-hoc effort
 - ▶ but unethical in high risk settings
- 2 “newer” approach
 - ▶ **indirectly** quantifies adult abundances
 - ▶ no *human landing collections* performed

The two approaches to estimate adult abundances

- 1 “gold standard” approach
 - ▶ based on human landing collections
 - ▶ directly quantifies adult abundances
 - ▶ but expensive and requires an ad-hoc effort
 - ▶ but unethical in high risk settings
- 2 “newer” approach
 - ▶ **indirectly** quantifies adult abundances
 - ▶ no *human landing collections* performed
 - ▶ leverages on existing monitoring systems

The two approaches to estimate adult abundances

- 1 “gold standard” approach
 - ▶ based on human landing collections
 - ▶ directly quantifies adult abundances
 - ▶ but expensive and requires an ad-hoc effort
 - ▶ but unethical in high risk settings
- 2 “newer” approach
 - ▶ **indirectly** quantifies adult abundances
 - ▶ no *human landing collections* performed
 - ▶ leverages on existing monitoring systems
 - ▶ abundance is measured as eggs abundance (in ovitraps)

The “newer” solution in practice

Relies on $f(\text{eggs}) = \text{Adult_Abundance}$

The “newer” solution in practice

Relies on $f(\text{eggs}) = \text{Adult_Abundance}$

- the R_0 formula requires adult abundance as an input

The “newer” solution in practice

Relies on $f(\text{eggs}) = \text{Adult_Abundance}$

- the R_0 formula requires adult abundance as an input
- this $f()$ may be as simple as a coefficient or... very complex

Question: How to implement the “newer” solution

Researchers of the “Gruppo Zanzare” at SUPSI¹ **collected some data** about adults and eggs abundances and wanted to know **how to proceed?**

¹University of Applied Sciences and Arts of Southern Switzerland, [Link to group homepage](#)

Question: How to implement the “newer” solution

Researchers of the “Gruppo Zanzare” at SUPSI¹ **collected some data** about adults and eggs abundances and wanted to know **how to proceed?**

- simple correlation?

¹University of Applied Sciences and Arts of Southern Switzerland, [Link to group homepage](#)

Question: How to implement the “newer” solution

Researchers of the “Gruppo Zanzare” at SUPSI¹ **collected some data** about adults and eggs abundances and wanted to know **how to proceed?**

- simple correlation?
- Generalised Least Squares?

¹University of Applied Sciences and Arts of Southern Switzerland, [Link to group homepage](#)

Question: How to implement the “newer” solution

Researchers of the “Gruppo Zanzare” at SUPSI¹ **collected some data** about adults and eggs abundances and wanted to know **how to proceed?**

- simple correlation?
- Generalised Least Squares?
- Generalised Mixed-Effects Models?

¹University of Applied Sciences and Arts of Southern Switzerland, [Link to group homepage](#)

Question: How to implement the “newer” solution

Researchers of the “Gruppo Zanzare” at SUPSI¹ **collected some data** about adults and eggs abundances and wanted to know **how to proceed?**

- simple correlation?
- Generalised Least Squares?
- Generalised Mixed-Effects Models?

What is our role?

¹University of Applied Sciences and Arts of Southern Switzerland, [Link to group homepage](#)

Question: How to implement the “newer” solution

Researchers of the “Gruppo Zanzare” at SUPSI¹ **collected some data** about adults and eggs abundances and wanted to know **how to proceed?**

- simple correlation?
- Generalised Least Squares?
- Generalised Mixed-Effects Models?

What is our role?

Suggest the best-suited method to use

¹University of Applied Sciences and Arts of Southern Switzerland, [Link to group homepage](#)

Question: Who is going to analyse the data

What is our role?

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Often the complexity of the analyses implies that is it better that we run the analyses.

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Often the complexity of the analyses implies that is it better that we run the analyses.

- the “obvious” reasons:

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Often the complexity of the analyses implies that is it better that we run the analyses.

- the “obvious” reasons:
 - ▶ quality of the analyses

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Often the complexity of the analyses implies that is it better that we run the analyses.

- the “obvious” reasons:
 - ▶ quality of the analyses
 - ▶ efficiency

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Often the complexity of the analyses implies that is it better that we run the analyses.

- the “obvious” reasons:
 - ▶ quality of the analyses
 - ▶ efficiency
- the “hidden” reasons:

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Often the complexity of the analyses implies that is it better that we run the analyses.

- the “obvious” reasons:
 - ▶ quality of the analyses
 - ▶ efficiency
- the “hidden” reasons:
 - ▶ model validation

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Often the complexity of the analyses implies that is it better that we run the analyses.

- the “obvious” reasons:
 - ▶ quality of the analyses
 - ▶ efficiency
- the “hidden” reasons:
 - ▶ model validation
 - ▶ model/method extrapolability

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Often the complexity of the analyses implies that is it better that we run the analyses.

- the “obvious” reasons:
 - ▶ quality of the analyses
 - ▶ efficiency
- the “hidden” reasons:
 - ▶ model validation
 - ▶ model/method extrapolability
 - ▶ new ideas, how to sell

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Often the complexity of the analyses implies that is it better that we run the analyses.

- the “obvious” reasons:
 - ▶ quality of the analyses
 - ▶ efficiency
- the “hidden” reasons:
 - ▶ model validation
 - ▶ model/method extrapolability
 - ▶ new ideas, how to sell
 - ▶ efficient publication process

Question: Who is going to analyse the data

What is our role?

Shall we run the analyses?

Often the complexity of the analyses implies that is it better that we run the analyses.

- the “obvious” reasons:
 - ▶ quality of the analyses
 - ▶ efficiency
- the “hidden” reasons:
 - ▶ model validation
 - ▶ model/method extrapolability
 - ▶ new ideas, how to sell
 - ▶ efficient publication process
 - ▶ ...

Question: Who is going to analyse the data

What is our role?

Question: Who is going to analyse the data

What is our role?

Shall we ALWAYS run the analyses?

Question: Who is going to analyse the data

What is our role?

Shall we ALWAYS run the analyses?

- No!

Question: Who is going to analyse the data

What is our role?

Shall we ALWAYS run the analyses?

- No!
- ideally a mix of the analyses run by the researchers and others run by ourselves

Question: Who is going to analyse the data

What is our role?

Shall we ALWAYS run the analyses?

- No!
- ideally a mix of the analyses run by the researchers and others run by ourselves
- support them with the analyses they run

Question: Who is going to analyse the data

What is our role?

Shall we ALWAYS run the analyses?

- No!
- ideally a mix of the analyses run by the researchers and others run by ourselves
- support them with the analyses they run
- have them on-board with the analyses we run

Question: Who is going to analyse the data

What is our role?

Shall we ALWAYS run the analyses?

- No!
 - ideally a mix of the analyses run by the researchers and others run by ourselves
 - support them with the analyses they run
 - have them on-board with the analyses we run
- ⇒ knowledge transfer

What is our role: Enhance their data-driven thinking

What is our current role?

What is our role: Enhance their data-driven thinking

What is our current role?

- **enhance their data-driven thinking**

What is our role: Enhance their data-driven thinking

What is our current role?

- **enhance their data-driven thinking**
- such that the full potential of data science can be exploited in scientific research

What is our role: Enhance their data-driven thinking

What is our current role?

- **enhance their data-driven thinking**
- such that the full potential of data science can be exploited in scientific research
- if the researchers don't get what we do...

What is our role: Enhance their data-driven thinking

What is our current role?

- **enhance their data-driven thinking**
- such that the full potential of data science can be exploited in scientific research
- if the researchers don't get what we do...
- then they won't be able to communicate these findings to the scientific community

What researchers can learn from us

What researchers can learn from us

- how to best collect data

What researchers can learn from us

- how to best collect data
- how to best design experiments

What researchers can learn from us

- how to best collect data
- how to best design experiments
- be aware of what is potentially feasible/not feasible

What researchers can learn from us

- how to best collect data
- how to best design experiments
- be aware of what is potentially feasible/not feasible
- right now many analyses are a compromise between what is appropriate and what can be successfully communicated

Tiger Mosquito Insecticide example

Benefits resulting from the knowledge-transfer of the previous collaboration:

Tiger Mosquito Insecticide example

Benefits resulting from the knowledge-transfer of the previous collaboration:

Example: Testing the efficacy of an insecticide

Tiger Mosquito Insecticide example

Benefits resulting from the knowledge-transfer of the previous collaboration:

Example: Testing the efficacy of an insecticide

- fully appropriate design
(addressing directly hypotheses)

Tiger Mosquito Insecticide example

Benefits resulting from the knowledge-transfer of the previous collaboration:

Example: Testing the efficacy of an insecticide

- fully appropriate design
(addressing directly hypotheses)
- better data coding

Tiger Mosquito Insecticide example

Benefits resulting from the knowledge-transfer of the previous collaboration:

Example: Testing the efficacy of an insecticide

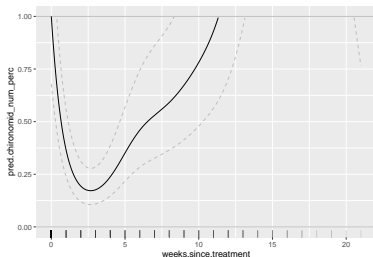
- fully appropriate design
(addressing directly hypotheses)
- better data coding
- better data structure

Tiger Mosquito Insecticide example

Benefits resulting from the knowledge-transfer of the previous collaboration:

Example: Testing the efficacy of an insecticide

- fully appropriate design (addressing directly hypotheses)
- better data coding
- better data structure
- Generalised Additive Models explained



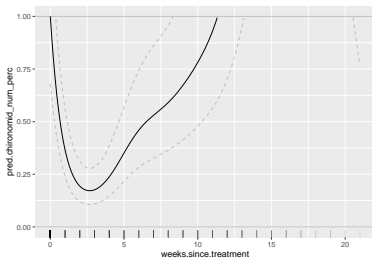
Tiger Mosquito Insecticide example

Benefits resulting from the knowledge-transfer of the previous collaboration:

Example: Testing the efficacy of an insecticide

- fully appropriate design (addressing directly hypotheses)
- better data coding
- better data structure
- Generalised Additive Models explained

⇒ this leads to a *win*³ situation



How to maximise (on-project) knowledge transfer

How to maximise (on-project) knowledge transfer

- invest a lot of energy in understanding what they really want to show

How to maximise (on-project) knowledge transfer

- invest a lot of energy in understanding what they really want to show
- invest a lot of energy in understanding what the underlying process is (ask many questions)

How to maximise (on-project) knowledge transfer

- invest a lot of energy in understanding what they really want to show
- invest a lot of energy in understanding what the underlying process is (ask many questions)
- get your hands dirty

How to maximise (on-project) knowledge transfer

- invest a lot of energy in understanding what they really want to show
- invest a lot of energy in understanding what the underlying process is (ask many questions)
- get your hands dirty
- frequent meetings

How to maximise (on-project) knowledge transfer

- invest a lot of energy in understanding what they really want to show
- invest a lot of energy in understanding what the underlying process is (ask many questions)
- get your hands dirty
- frequent meetings
- nicely commented, reproducible and easy to follow analyses (via dynamic documents)

How to maximise (on-project) knowledge transfer

- invest a lot of energy in understanding what they really want to show
- invest a lot of energy in understanding what the underlying process is (ask many questions)
- get your hands dirty
- frequent meetings
- nicely commented, reproducible and easy to follow analyses (via dynamic documents)
- discuss results in depth (focus on graphs)

Section 4

Conclusions

Conclusions

- data science represents a huge potential for scientific research

Conclusions

- data science represents a huge potential for scientific research
- researchers need active help for the transition to data-driven thinking research

Conclusions

- data science represents a huge potential for scientific research
- researchers need active help for the transition to data-driven thinking research
- data scientists supporting researchers carry the responsibility of make this transition possible

Conclusions

- data science represents a huge potential for scientific research
- researchers need active help for the transition to data-driven thinking research
- data scientists supporting researchers carry the responsibility of make this transition possible
- the transition to data-driven thinking research is interest of all actors involved

End

Looking forward to your questions, comments, and inputs!

Section 5

Additional slides

How to maximise (formal) knowledge transfer (ie. Teaching)

I wonder...

How to maximise (formal) knowledge transfer (ie. Teaching)

I wonder...

- via formal courses (teaching)

How to maximise (formal) knowledge transfer (ie. Teaching)

I wonder...

- via formal courses (teaching)
- via “less formal” courses (hands-on teaching)

How to maximise (formal) knowledge transfer (ie. Teaching)

I wonder...

- via formal courses (teaching)
- via “less formal” courses (hands-on teaching)
- via new ways of teaching

How to maximise (formal) knowledge transfer (ie. Teaching)

I wonder...

- via formal courses (teaching)
- via “less formal” courses (hands-on teaching)
- via new ways of teaching
- via (on the project) knowledge transfer

How to maximise (formal) knowledge transfer (ie. Teaching)

I wonder...

- via formal courses (teaching)
- via “less formal” courses (hands-on teaching)
- via new ways of teaching
- via (on the project) knowledge transfer
- via inclusion during all steps of any collaboration project

How to maximise (formal) knowledge transfer (ie. Teaching)

I wonder...

- via formal courses (teaching)
- via “less formal” courses (hands-on teaching)
- via new ways of teaching
- via (on the project) knowledge transfer
- via inclusion during all steps of any collaboration project
- via new targets of teaching (group leaders?)

Win-win: What are *our* benefits

Win-win: What are *our* benefits

- we can still be involved in research which we love

Win-win: What are *our* benefits

- we can still be involved in research which we love
- ... but we don't need to bother about looking for the next post-doc

Win-win: What are *our* benefits

- we can still be involved in research which we love
- ... but we don't need to bother about looking for the next post-doc
- we can help creating better science

Win-win: What are *our* benefits

- we can still be involved in research which we love
- ... but we don't need to bother about looking for the next post-doc
- we can help creating better science
- benefits for a wider range of people

Win-win: What are *our* benefits

- we can still be involved in research which we love
- ... but we don't need to bother about looking for the next post-doc
- we can help creating better science
- benefits for a wider range of people
- we can help create more transparent/reproducible science

Win-win: What are *our* benefits

- we can still be involved in research which we love
- ... but we don't need to bother about looking for the next post-doc
- we can help creating better science
- benefits for a wider range of people
- we can help create more transparent/reproducible science
- which in turn is expected to improve science perception

Win-win: What are *our* benefits

- we can still be involved in research which we love
- ... but we don't need to bother about looking for the next post-doc
- we can help creating better science
- benefits for a wider range of people
- we can help create more transparent/reproducible science
- which in turn is expected to improve science perception
- to get to the next level of data science (AI) projects

Win-win: What are *our* benefits

- we can still be involved in research which we love
- ... but we don't need to bother about looking for the next post-doc
- we can help creating better science
- benefits for a wider range of people
- we can help create more transparent/reproducible science
- which in turn is expected to improve science perception
- to get to the next level of data science (AI) projects
- to see the full potential of data science exploited