

A FANCY UNIVERSITY

AN EVEN Fancier LAB

*Christian Stentoft Michelsen*

tufte-style-thesis,  
a Tufte-styled L<sup>A</sup>T<sub>E</sub>X class for  
theses

Actually more of a mix  
between Edward Tufte and Robert Bringhurst

Doctoral thesis

*November 9, 2022*

Supervisor	their name	their job
Cosupervisor	also their name	also their job
Jury members	jury 1	jobs
	jury 2	...

Christian Stentoft Michelsen, *tufte-style-thesis*,  
*a Tufte-styled L<sup>A</sup>T<sub>E</sub>X class for theses*, Actually more of a mix  
between Edward Tufte and Robert Bringhurst, November 9, 2022.

*Til de to kvinder i mit liv*



# *Contents*

Preface	7
Acknowledgements	9
Abstract	11
Dansk Abstract	13
Publications	15
1 Notes on the design	17
1.1 Document layout	17
2 Ancient DNA	19
3 Hospital and Machine Learning	21
4 COVID-19 and Agent Based Modelling	23
5 Bayesian Inference and Diffusion	25



# *Preface*

This Ph.D. thesis summarizes my scientific research in collaboration with the Niels Bohr Institute (NBI) and the Globe Institute, University of Copenhagen, and was funded by the Lundbeck Foundation. The research was supervised by Associate Professor Troels C. Petersen (NBI) and Assistant Professor Thorfinn S. Korneliussen (Globe Institute).

Being a multi-disciplinary project, the research presented in this thesis is multi-faceted and covers a wide range of topics with the main scope being the development and integration of a novel statistical methods and machine learning models for the analysis of large-scale biological data. The thesis is organized as follows. First I present a brief introduction to the statistical methods and machine learning models used in the thesis. Then I present the research in the form of four papers, each of which reflects a different aspect of the research.

The first paper presents a novel method I developed for detecting and classifying ancient DNA damage in metagenomic samples taking the full taxonomic information into account. While the first paper focuses on the development of the statistical model in the field of ancient genomics, the second paper focuses on the use of modern machine learning models in medicine and how advanced boosted decision trees can not only improve the accuracy of identifying patients at risk of being readmitted after knee or hip surgery, but doing so in a way that is interpretable as well. In the beginning of 2020 we all experienced how COVID-19 suddenly changed our lives and impacted our societies in dramatic ways. During this time, I ended up working for Statens Serum Institut (SSI), the Danish CDC, on a project to develop an agent based model capable of simulating the spread of COVID-19 in Denmark. This model is presented in the third paper and was used to inform the Danish government on how to best handle the pandemic in the early stages and the effect of lockdowns and other measures. Finally, in the fourth paper we show how advanced Bayesian methods can be utilized to better estimate the diffusion coefficients of molecules in the cell nucleus in XXX experiments.





# *Acknowledgements*

First of all I want to express my sincere gratitude to my long time supervisor, captain, and friend: Troels. You are truly an inspiration to work with. I want to thank you for opening so many doors for me, both academically, professionally, and nautically. I am looking forward to our future adventures together. I also want to thank my co-supervisor, Thorfinn. I want to thank you for introducing me to the field of bioinformatic and helping me to develop my skills in this area. I also want to thank you for your patience and guiding me through the endless amount of (near) identical biological concepts and helping me to understand the minute differences.

I have been fortunate to work with people from a wide range of backgrounds and disciplines during my Ph.D. The author lists on the papers in this thesis include a particle physicist, bioinformaticians, a clinical professor, epidemiologists, a medical doctor, a bio-physicist, a mathematician, a biologist, and the president of the Royal Danish Academy of sciences and letters. Before anything else, I want to thank all of my co-authors for their work and contributions to these papers and for allowing me to be a part of their projects. I have learned a lot from all of you, and I hope that I have been able to contribute something to your work as well.

I am thankful for all the people who have helped me with my work and listened to my complaints when I was stuck, when the code did not compile, or when the small bug was almost impossible to find (which was not a small amount of time). In particular I want to thank the people at Globe who I have spent the most time with; Rasa, Alba, and Rasmus. I also want to thank the Korneliussen Group and the people in my office for helpful advice, suggestions and discussions. This also includes Daniel Nielsen and Rasmus Ørsøe from NBI. Finally, I want to thank Mathias Heltberg for many years of fruitful collaboration and for including me in his projects.

This project would not have been possible had it not been for the Lundbeck Foundation which funded my Ph.D. In addition to the funding itself, I am grateful for the inter-disciplinary aspect of project which has allowed me to meet so many inspiring and talented people and for the freedom to pursue my own interests within the project.

I would also like to express my gratitude to Professor Guido Sanguinetti from the International School for Advanced Studies, SISSA, in Trieste, Italy, for hosting me in his group during the Winter of 2021/2022. My gratitude also goes out to Kosio, Sara, Max, Romina, Noor, Viplove, Anne-Marie, and all the other wonderful people that I met during in Trieste. Thanks for making my stay in Italy so enjoyable and for welcoming me in a way that only non-Danes can do.

I want to thank my friends for always being there for me. A special thanks to my friends from NBI and Borchsen who I know that I can always count on, whether or not that includes a trip in the party bus of the Sea, taking Artemis out for a sail, or board games and beer. Thank you for always being there. I also want to thank my family, especially my parents for their unconditional support and encouragement. I am grateful for the opportunities that they have given me and for the sacrifices that they have made for me.

Lastly, I want to thank my future wife and mother of our child, Anna. I would not have been able to do this without you. Thank you for your patience and support. I am looking forward to our future together. I love you.

# *Abstract*

Basically a thesis (book?) class for Tufte lovers like myself. I am aware that `tufte-latex` already exists but I just wanted to create my own thing.



## *Dansk Abstract*

Her et dansk abstract.



## *Publications*

Here a list of publications.





# 1 *Notes on the design*

This class is my personal mix of different book design influences: mainly the works of Edward R. Tufte,<sup>1-4</sup> known for the big margin and the plentiness of sidenotes and sidecaptions. The margins are however not as prominent as in Tufte's works, the main text takes a bit more space, more like in Robert Bringhurst's typographer's bible [5]. So it is a bit of a mix of Tufte and Bringhurst, with some of my own choices for other design features, as we will see through this chapter.

## 1.1 *Document layout*

While `tufte-style-thesis` is a class for typesetting theses, the general layout is pretty much the same as in a regular book. A book is traditionally divided into three major sections: the front matter, the main matter and the back matter.

<sup>1</sup>Tufte, E. R. *The Visual Display of Quantitative Information* ISBN: 0-9613921-4-2 (Graphics Press, Cheshire, Connecticut, 2001).

<sup>2</sup>Tufte, E. R. *Beautiful Evidence* ISBN: 0-9613921-7-7 (Graphics Press, Cheshire, Connecticut, 2006).

<sup>3</sup>Tufte, E. R. *Envisioning Information* ISBN: 0-9613921-1-8 (Graphics Press, Cheshire, Connecticut, 1990).

<sup>4</sup>Tufte, E. R. *Visual Explanations* ISBN: 0-9613921-2-6 (Graphics Press, Cheshire, Connecticut, 1997).



## **2**    *Ancient DNA*



### **3**    *Hospital and Machine Learning*



## 4 *COVID-19 and Agent Based Modelling*





## 5 *Bayesian Inference and Diffusion*



## *Bibliography*

1. Tufte, E. R. *The Visual Display of Quantitative Information* ISBN: 0-9613921-4-2 (Graphics Press, Cheshire, Connecticut, 2001).
2. Tufte, E. R. *Beautiful Evidence* ISBN: 0-9613921-7-7 (Graphics Press, Cheshire, Connecticut, 2006).
3. Tufte, E. R. *Envisioning Information* ISBN: 0-9613921-1-8 (Graphics Press, Cheshire, Connecticut, 1990).
4. Tufte, E. R. *Visual Explanations* ISBN: 0-9613921-2-6 (Graphics Press, Cheshire, Connecticut, 1997).
5. Bringhurst, R. *The Elements of Typographic Style* ISBN: 978-0881791327 (Hartley & Marks Publishers, 2002).



This document was typeset using  $\LaTeX$  and the `tufte-style-thesis` class.  
The style is heavily inspired by the works of Edward R. Tufte and Robert Bringhurst.  
This is available on here:

<https://github.com/sylvain-kern/tufte-style-thesis/>.

Feel free to contribute!