# Structure of my thesis (Draft 1.3)

* Connections unclear
* Not sure what to focus on

# Abstract (~300)

* Motivation (30)
* Problem statement (60)
* How I approached the problem(60)
* Results (60)
* Conclusions (90)
* Key words \* 10

# Back ground (~5000)

* **Introduction (300)**
  + History and brief info about the topic, disease surveillance?
* **Motivation (600**)
  + Why hierarchical model is useful
  + Why we interested in anomaly detection in hierarchical data
  + Problems with hierarchical data and
  + What has people done/looked into
* **Hierarchical model (800)**
  + Matrix
  + Equations and Notations
  + Algorithms
* **Bayesian Model (800)**
  + How Bayesian is applied
  + Why it is useful
  + Poisson distributions
  + How this suit our data
  + Algorithms
* **Anomaly detection (800)**
  + Intro
  + Types of anomaly
  + Detection algorithms
* **Outline of Thesis (800)**
  + What did I implement
  + What kind result is expected
  + Chapter summary (1 Sentence /chapter)
* **Notational Conventions Used (~200)**
  + Summary of different custom notation/font used in thesis
  + R-codes, Packages, web links, R functions, SAS Codes

# Data (~3000)

* Collection of Data (400)
  + What kind of data usually have hierarchical structure
  + Where can we find our data
* ICD 10 cm Intro (400)
* Structure of ICD 10cm (1200)
* Anomalies we expect / detection requirements (600)
* Discussion (300)
* Conclusion (100)

# Algorithms and proofs (~4000)

Mathematical proofs for algorithms used

* Hierarchical structure
* Bayesian models
* Anomaly detection
* Error/reliability of algorithms

# Efficiency (~3000)

* Problems with large dataset (200) what I should use for faster performance and what is the cost
* BUGS (600)
* JAG (600)
* STAN (600)
* Test with Small Dataset and evaluate efficiency (800)
* Discussion (300)
* Conclusion (100)

# Simulated results (~6000)

* Create data (500)
* Test alpha and beta for anomaly (1500)
* Expand on idea and look at other stuff, ie type, size, amount etc, (3000)
* Discussion (600)
* Conclusion (200)

# Analysis & other statistics (~4000)

* Discussion (300)
* Conclusion (100)

# ED Data and Results ~(6000)

* Intro of data
* Methodology
* Explanation of graphs
* Summary of results
* Conclude results
* Specific case study
* Conclusion

# Discussion (~4000)

* Explanation of results (800)
* Patterns found
* Expected and Unexpected finds
* References to previous research, how consistent is our result
* How the results can be applied more generally.
* What conclusions can be made

# Conclusion (~2000)

* Summary of result (600)
* Deduction of result (300)
* Personal opinion (300)
  + To what extent did I achieve my aim
  + Raise question on statistical and practical significance
* Implications (300)
* Limitations (300)
* **Future Directions (300)**
  + What your research demonstrate, what could you improve on
  + Interesting questions raised
  + suggest something that could be developed from your work as a PhD thesis

(38,000)

# (Coding~2000-4000)

* Use Knit R to output and Insert into PDF later?
* Simulated data
* ED Data
  + Data cleaning
  + Simulation
  + Analysis
* Other stuff

# (Reference~3000)

* Need probably 80-120

# Extra stuff, If there is time to spare

* Practise on end of year presentation (2 week)
* Make a 1 hr Presentation on tape (1 week)
* Make a poster (1 day)
* Write a Chinese Version for bilingual understanding(1 Week)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | Weeks |  |  |  |  |  |  |  |  |  |
|  |  |  | 1 |  |  | Literature Reviews | |  | Notes |  |  |  |
|  |  |  | 2 |  |  |  |  |  | \* Keep Thomas updated | | |  |
|  |  |  | 3 |  |  |  |  |  | \* Structure everything | | |  |
|  |  |  | 4 |  |  |  |  |  | \* Always take longer than expected | | | |
|  |  |  | 5 |  |  |  |  |  |  |  |  |  |
|  |  |  | 6 |  |  |  |  |  |  |  |  |  |
|  |  |  | 7 |  |  |  |  |  |  |  |  |  |
|  |  |  | 8 |  |  |  |  |  |  |  |  |  |
|  |  |  | 9 |  |  | Proposal |  |  |  |  |  |  |
|  |  |  | 10 |  |  |  |  |  |  |  |  |  |
|  |  |  | 11 |  |  |  |  |  |  |  |  |  |
| 10-Oct |  |  | **12** |  |  |  |  |  |  |  |  |  |
|  |  |  | 13 |  |  | Background | |  |  |  |  |  |
|  |  |  | 14 |  |  | Data |  |  |  |  |  |  |
|  |  |  | 15 |  |  | Algorithms | |  |  |  |  |  |
|  |  |  | 16 |  |  |  |  |  |  |  |  |  |
|  |  |  | 17 |  |  | 782 Exm |  |  |  |  |  |  |
|  |  |  | 18 |  |  |  |  |  |  |  |  |  |
|  |  |  | 19 |  |  | Methodology | | \* BUGS |  |  |  |  |
|  |  |  | 20 |  |  |  |  | \*JAGS |  |  |  |  |
|  |  |  | 21 |  |  |  |  | \*STAN |  |  |  |  |
|  |  |  | 22 |  |  |  |  |  |  |  |  |  |
|  |  |  | 23 |  |  |  |  |  |  |  |  |  |
| 24-Dec |  |  | **24** |  |  |  |  |  |  |  |  |  |
|  |  |  | 25 |  |  | Pilot test/Simulation | | |  |  |  |  |
|  |  |  | 26 |  |  |  |  |  |  |  |  |  |
|  |  |  | 27 |  |  |  |  |  |  |  |  |  |
|  |  |  | 28 |  |  | Main test/ ED Data | |  |  |  |  |  |
|  |  |  | 29 |  |  |  |  |  |  |  |  |  |
|  |  |  | 30 |  |  |  |  |  |  |  |  |  |
|  |  |  | 31 |  |  |  |  |  |  |  |  |  |
|  |  |  | 32 |  |  |  |  |  |  |  |  |  |
|  |  |  | 33 |  |  |  |  |  |  |  |  |  |
|  |  |  | 34 |  |  |  |  |  |  |  |  |  |
|  |  |  | 35 |  |  |  |  |  |  |  |  |  |
| 18-Mar |  |  | **36** |  |  |  |  |  |  |  |  |  |
|  |  |  | 37 |  |  | LateX Report | |  |  |  |  |  |
|  |  |  | 38 |  |  |  |  |  |  |  |  |  |
|  |  |  | 39 |  |  |  |  |  |  |  |  |  |
|  |  |  | 40 |  |  |  |  |  |  |  |  |  |
|  |  |  | 41 |  |  |  |  |  |  |  |  |  |
|  |  |  | 42 |  |  |  |  |  |  |  |  |  |
|  |  |  | 43 |  |  | Touch up on english, latex structure etc | | | |  |  |  |
|  |  |  | 44 |  |  |  |  |  |  |  |  |  |
|  |  |  | 45 |  |  | ppt, Poster, maybe write a chinese translation | | | | |  |  |
|  |  |  | 46 |  |  |  |  |  |  |  |  |  |
|  |  |  | 47 |  |  | Practise for PPT | |  |  |  |  |  |
| 17-Jun |  |  | **48** |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | | |  |  |  |  |