

!n f er → en t ial°

!n t er → es t s₊



INSERT
CALCULUS
TO BEGIN

TABLE OF CONTENTS

3 LETTER FROM THE EDITORS

4 CONFERENCE INTERVIEW WITH MINH-NGOC

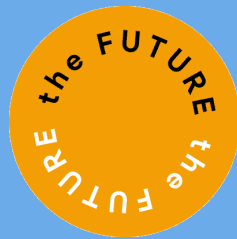
9 FIVE TIPS: STATS CONSULTING

11 INTERVIEW WITH EMI

15 REVIEWS: MUSIC AND BOOKS

20 MY JOURNEY: CLOSE CONTACT TRACING

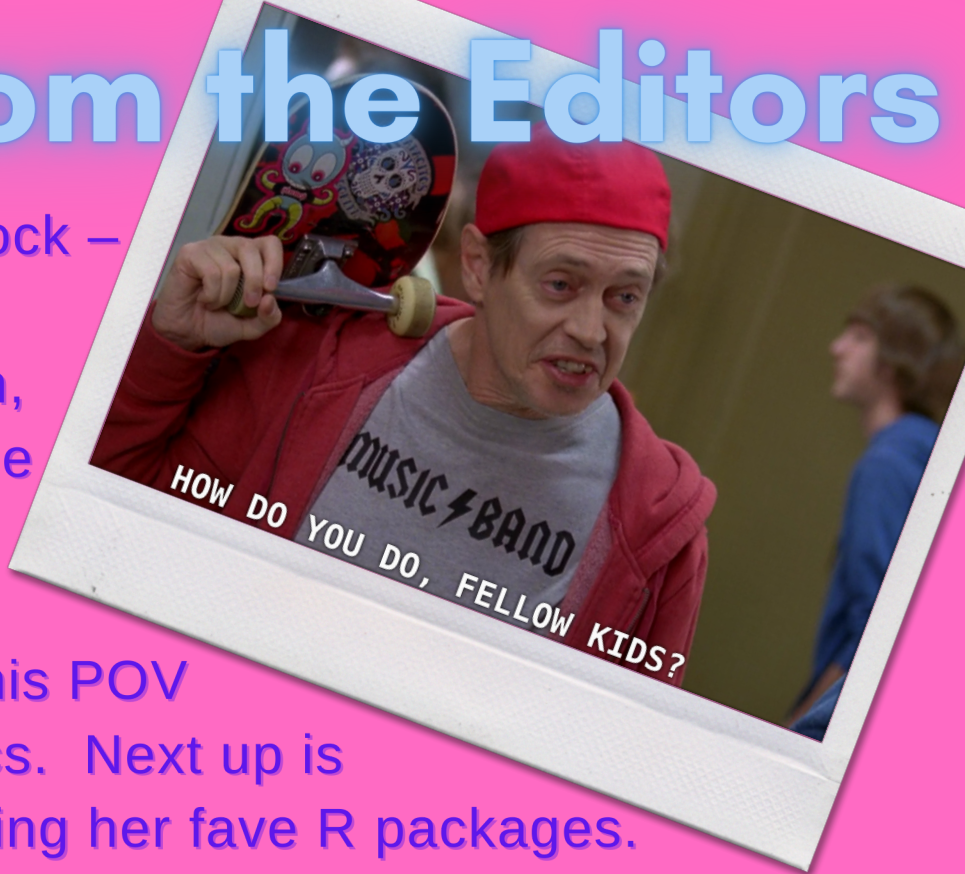
22 SPONSORS



STATISTICS LECTURER
DR.EMI TANAKA
HAS THE PACKAGES
YOU NEED

Letter from the Editors

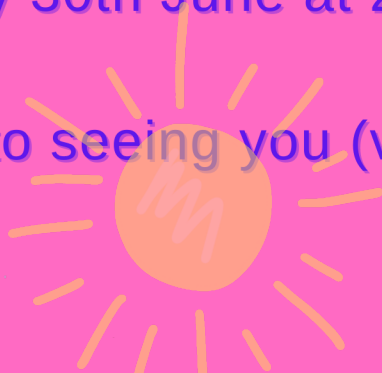
This issue is chocka block – check out the interview with Dr Minh-Ngoc Tran, a keynote speaker at the Early Career and Student Statisticians Conference, giving us his POV on the future of statistics. Next up is Dr Emi Tanaka discussing her fave R packages. Our own Cameron, gives us his top tips for statistical consulting, and *library(books)* looks at handy texts for those Stata users out there. Two more of our committee members write in with their stories - Ben on getting a black eye from djent music and Jordan on contact tracing.



Check out our podcast with Peter Taylor at *The Random Sample!* Thank you for having us ACEMS - we loved it. Also don't forget to register for the ECSSC! Our team is busy behind the scenes putting together a totally awesome virtual satchel for you, including lotsa vouchers. Also, next up in the ECSSC Science Communication Series is *Presenting Virtually* with Dr Karen Lamb, which will take place Wednesday 30th June at 2 pm!

Looking forward to seeing you (virtually) at the conference,

CAT AND SPLIT





MINH-NGOC TRAN

Interviewed by Catriona Croton

One of our keynote speakers, Minh-Ngoc Tran, joins us to talk about his research, the future of statistics and... coffee. Minh-Ngoc is an Associate Professor at the Business Analytics discipline, University of Sydney Business School, and an Associate Investigator at ACEMS. His research interests are Bayesian methodology and statistical machine learning, and he co-wrote an R package deepglm.

UPCOMING CONFERENCE

ON HIS FIRST JOB IN THE FIELD:

After my undergraduate degree in mathematics at Vietnam National University in Hanoi, I spent six months working with a team developing credit scoring systems for a bank in Vietnam. It was fun, but it required statistics, which I didn't know much about.

So I quit the job and went back to university to do a Masters in mathematics, majoring in statistics – at the time at VNU you couldn't study an entire degree in statistics at an undergraduate or Masters level. Next, I went to the National University of Singapore to do a PhD in statistics, and then straight onto the Business School at the University of New South Wales as a post-doctoral research fellow, in which I worked on Bayesian computation. After a few years, this job at University of Sydney came up at the Business School and I applied. I wouldn't say I planned to have a career in statistics in business – the jobs came up in that field and so I applied.



ON A TYPICAL WORKDAY:

In a typical day, the morning is spent on teaching and admin, and I do research in the afternoon. I am an afternoon person; that is when I focus. My formal time allocation is 40% Teaching, 40% Research, 20% admin, but I am more of a researcher guy.

COVID hasn't changed my workday much, although now I work at home most of the time. I think working from home is more efficient - I can catch up with new research, review papers, talk to research students on Zoom. And drink coffee. I am a heavy coffee drinker. Morning, afternoon and night. And I still get the usual seven hours of sleep a night. When you drink enough coffee it doesn't affect your sleep at all (laughs).



ON PROGRAMMING LANGUAGES

I use Matlab and Python. Because my school decided to pick up Python in teaching, I have done so too, but my favourite is Matlab. Everyone in my group can use Matlab, so we can learn from each other.

A NEW WAY OF WORKING:

In the last few years I have worked with collaborators outside statistics more frequently, like psychology, marketing and finance, and I realised that I learned a lot from them. They are not statisticians, but the way they explain or communicate the statistical findings is interesting. We as statisticians can learn from them.

FUTURE OF STATISTICS/ DATA SCIENCE:

My feeling is we are moving towards automated statistics and data science, for the general public. Today everyone uses computers, even though most don't know how the computer actually works. One day when data is everywhere, and it is convenient and easy to store this data, everyone will want to do data analysis on their own in an automated manner, like they use a computer today.



HOW TO PREPARE FOR CHANGES IN THE PROFESSION:

I think the theory in statistics is still important – in order to be good at statistics, you need to be good at the theory. However, I think statistical computing is getting more and more important as you need this to bring statistics closer and closer to everyday life application.

I always say to my PhD students that you need to be good at three things to be a good statistician or data scientist:

- 1) Background in mathematics
- 2) Computation and programming skills
- 3) Writing and speaking skills to communicate to a general audience, not only experts



TIPS FOR WORK-FAMILY BALANCE:

I have none (laughs). I make sure I spend time on Saturday and Sunday with my family – take the kids to the park and so on. That is their time. Family is the most important thing. Finally, I just wanted to say I am looking forward to the conference!

THANK YOU MINH-NGOC FOR THE INTERVIEW,
AND WE ARE ALSO LOOKING FORWARD TO
SEEING YOU AT THE CONFERENCE!

5 TIPS: STATISTICAL CONSULTING

BY CAMERON PATRICK, STATISTICAL CONSULTING CENTRE,
UNIVERSITY OF MELBOURNE

Providing a genuinely useful service to the client requires a good understanding of their needs and research question, which only comes from asking a lot of questions. Don't be afraid to request an "explain like I'm five" description of any part of the study that doesn't make sense to you. Clients are experts in their own fields, and consulting is a great opportunity to learn from them while working on a collaborative solution. Enjoy this process - remember "being a statistical consultant means you have the fun of playing in everyone else's back yard" (Tukey).

1

STAY
CURIOUS

IT'S NOT
JUST
ABOUT
THE
STATS

2

You may have an advanced degree in statistics, but don't forget the other skills you'll need: report writing, project management, and juggling timelines to keep a bunch of clients happy. Being able to communicate clearly with your client, and explain statistical concepts in terms they are familiar with, is far more important than knowing the latest machine learning models.

3 THINK LIKE A COMPUTER SCIENTIST

Modern applied statistics is heavily computational, so being proficient in a software package like R or Stata is a huge advantage. It is a time saver in the long run to learn simple programming in your statistical software of choice to automate repetitive tasks and provide a reproducible workflow. It's worthwhile spending an afternoon learning to automate a process that would take an hour to do by hand, because you'll use that code again. Clients are increasingly asking for copies of the source code for their analysis; think of this as part of your finished product, not a throw-away intermediate step. I like to use a tool like R Markdown, which allows you to integrate the code into the report.

The human side of statistical consulting is as important as the statistical side. Often when clients engage a statistician, they're scared, intimidated, or frustrated by their prior experiences. Helping a client past that fear into a place where they're comfortable with the statistical concepts is rewarding for both consultant and client. Also, happy clients will come back again and again, and refer their colleagues your way. With time, you may form longer term collaborations with research groups who mesh well with your statistical skills and interests.

4 BUT STILL BE A FRIEND

Statisticians embedded in a group of non-statisticians often feel isolated from the rest of the statistical community. Beat that isolation by joining a shared conversation with communities of researchers and statisticians on social media like Twitter – it's a great way to share tips, learn from the experiences of others, and make friends. Organisations like the Statistical Society of Australia and the Early Career and Student Statisticians Network allow you to meet colleagues doing similar work, in your geographic area or virtually. So many conferences are fully or partially virtual now, that it is easier than ever to attend. Don't forget the ECSSC from 26th July to 1st August 2021!

5 STAY CONNECTED

MY FAVE PROGRAM: R



Dr. Emi Tanaka chose to do statistics from the start – straight from a Bachelor degree majoring in mathematics and statistics, to Honours in statistics, and immediately onto her PhD. Now working as lecturer in statistics at Monash University, her research includes combining elements of classical statistics with software development to create tools that can support scientists plan and design their experiments.

Transcribed by Fennic Oost
Written by Splithoof Rivera
Edited by Catriona Croton

WERE YOU ALWAYS IN STATISTICS OR DID YOU MOVE ACROSS FROM ANOTHER AREA?

I did do statistics from the beginning, but don't feel like you need to. I've met a lot of people who didn't do that and are equally successful.

WHY CHOOSE TO SPECIALISE IN R?

I learned R because it was what I was taught but I was using Python mainly in my PhD. I came back to using R because a lot of statistical modelling is much more advanced in R. It's really driven by my needs - when I need to do complex modelling, packages in other languages just aren't as sophisticated and well developed as R.

WHAT ARE YOUR FAVOURITE PACKAGES?

It's so hard to pick a favourite! I think I do like a fair bit of tidyverse which has the packages dplyr, tidyR and purrr for functional math programming. Also ggplot2 is great for visualising data. I also like some really specific packages like skimr which just allows you to get a summary of your data. Also there's another package called janitor - you know how when you put in the data it has really messy headers? This one function cleans all your headers! It's just one little thing, but I love it for that.

WHAT IS THE BEST ENVIRONMENT TO LEARN R?

First, it always helps to be motivated – for example, I use it to do your own budgeting and finances. I was looking at real estate to buy a property and I wrote code that harvests real estate prices to get an idea of what I was purchasing! Don't work on a hypothetical exercise because it's just not as exciting as working on something that's directly relevant to you.

WHERE DO YOU SEE R IN THE FUTURE?

I sometimes wonder because languages can disappear. However, there are entities like RStudio who have invested a lot into R and there has been considerable volunteer contribution - I think it will be a long time before R disappears.

WERE YOU GOOD AT MATHS FROM A YOUNG AGE?

I didn't think I was good at it; to this day I don't think I'm good at it. My marks were always good and I kept asking, '*Are you sure?*' You don't have to be smart to get it so long as you make contributions that are beneficial for others. That's what people really want when employing you.

DID YOU ENJOY STATS IMMEDIATELY?

Actually, I found statistics to be uninspiring at first. I was confused and it was '*oh histogram, oh bar plot, hypothesis testing*' and I just went through the motions. I think some people were much smarter than me at it, so I got my PhD instead (*laughs*). I'm much better at it now, which I think comes from practice. I didn't get it from the beginning; there was a long period of perseverance and struggle before I got to the point of '*I really like statistics! I get it now!*'

YOU MENTIONED THINKING YOU WEREN'T VERY SMART. MANY STUDENTS STRUGGLE WITH IMPOSTER SYNDROME.


I'm not quite sure that is how I felt - the term 'Imposter Syndrome' implies that you don't belong.

I never felt that clever, but I didn't feel I lacked belonging. When I say I'm not smart, I'm not putting myself down. I am acknowledging there are always people out there better than me. I wasn't born a genius, but I can contribute. There are things I can focus on and I can do. So I concentrated on those things and built upon that. I think to deter negative feelings, focus on what you can control.

THANK YOU EMIL FOR THE INTERVIEW!

Vildhjarta

Everybody has heard of djent... Just kidding, nobody has heard of djent. It's a music genre with Vildhjarta among the ranks. The name djent is onomatopoeic, from its characteristic guitar sound. A key component is your brain telling your body to mosh, but your body being unable figure out the rhythm. This is thanks to the polyrhythmic drumming and guitarwork, and which I personally thank for the wayward elbow that gave me a black eye in a Meshuggah mosh pit many years ago.



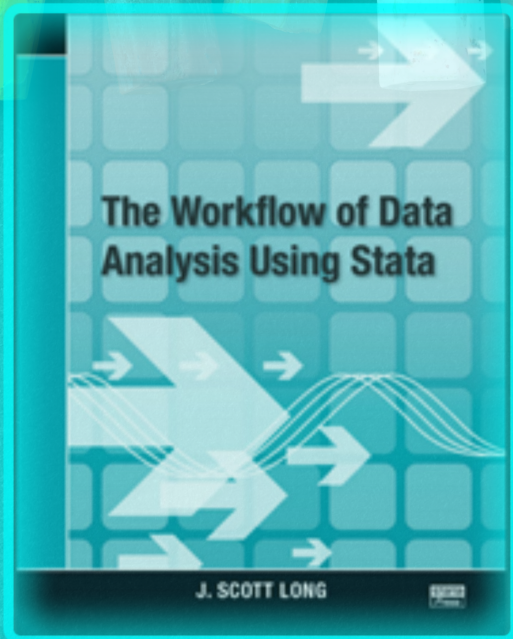
But I'm not here to talk about Meshuggah, I'm here to tell you about Vildhjarta. Heralding from Sweden, Vildhjarta, which translates as 'wild heart', are a standout but lesser-known djent band. Their debut album *Måsstaden* was released in 2011. The first song on the album *Shadow* begins innocently with some gentle, foreboding guitar, before getting down to business and brutalising your ears with ripping polyrhythmic riffs. From here on out the album doesn't slow down.

Listening to the complexity of the riffs and failing miserably to air-drum to the ever-changing tempo reminds me of spending hours working on a complicated mathematical proof for an assignment; thinking I'd got it then discovering I misread a letter and have to re-evaluate everything I'd done.

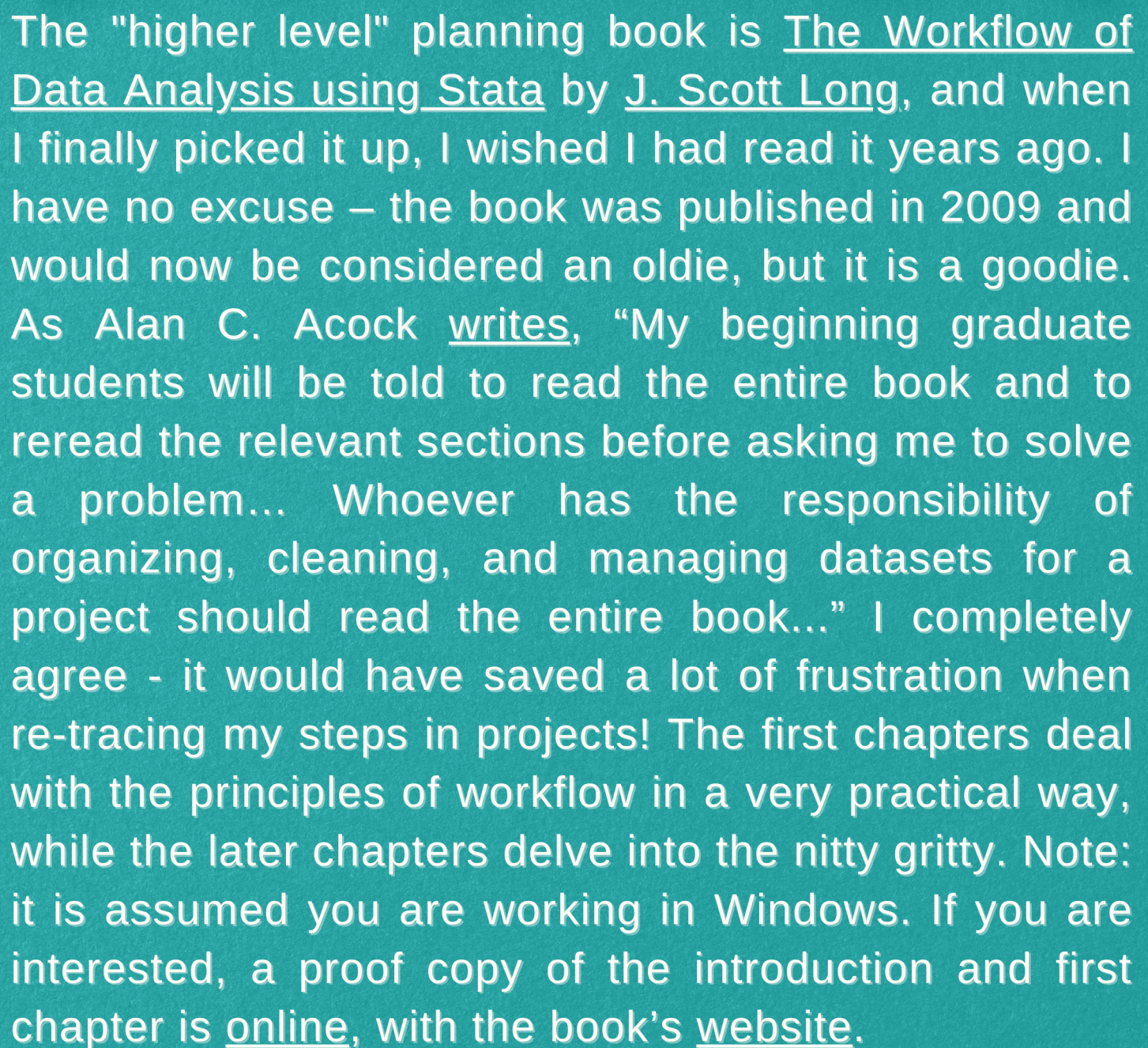
Roughly halfway through the album is my personal highlight - *Traces*. The mixing together of clean vocals with the familiar death metal growl gives me goosebumps. This album, played at unhealthy volumes, makes for excellent coding music. Their recent release of the single *Den Helige Anden* teases an album to hopefully be released this year, so keep your ears diced.

LIBRARY(BOOKS)

by Catriona Croton



This issue we have a couple of books that are universally helpful when working in Stata, no matter the project. They are a natural pair: the first book is for workflow to plan your projects – higher level planning – and the second is the data management – lower level operational. I usually hate the terms "higher level" and "lower level", as it implies the actual doing of the work is somehow less, but the terms do apply here. If you are like me, you do both kinds of work. Or should be!



The "higher level" planning book is The Workflow of Data Analysis using Stata by J. Scott Long, and when I finally picked it up, I wished I had read it years ago. I have no excuse – the book was published in 2009 and would now be considered an oldie, but it is a goodie. As Alan C. Acock writes, “My beginning graduate students will be told to read the entire book and to reread the relevant sections before asking me to solve a problem... Whoever has the responsibility of organizing, cleaning, and managing datasets for a project should read the entire book...” I completely agree - it would have saved a lot of frustration when re-tracing my steps in projects! The first chapters deal with the principles of workflow in a very practical way, while the later chapters delve into the nitty gritty. Note: it is assumed you are working in Windows. If you are interested, a proof copy of the introduction and first chapter is online, with the book’s website.

The “lower level” book is Data Management Using Stata by Michael N. Mitchell, with the second edition now out. This book deals with data management and manipulation – an often over-looked integral step before analysis – and takes you through a wide variety of tasks including data cleaning, using subgroups and looping. My favourite chapters are to do with reshaping and combining datasets with appending and merging – this book sits on my desk when I am squashing a dataset into a useable form as it has the clearest explanations I have seen on this.

If Stata is your program of choice, then these books are well worth your time - look for them in the library!



NSW CLOSE CONTACT TRACING EXPERIENCE

By [Jordan Hedi](#)

During the first half of 2020 I was working as a conduct risk analyst for Westpac, when I was called out of the blue by someone from New South Wales Health Recruitment asking if I would be interested in joining their Close Contact Tracing Team (CCTT).

I queried as to how they got my details, and was told someone from within the SSA's NSW branch who worked with NSW Health had passed them on. After that reassurance, I accepted the offer as I thought it would be a valuable way to assist in combating the pandemic.

The days leading up to my start with CCTT, I was hit by a paroxysm of doubt. But, luckily, I was put under the tutelage of Fiona Horn, who took me under her wing and gave me the skills and confidence to operate effectively on the phones. The calls would vary greatly - usually people were calm and understanding about the need to isolate. Other people would become very abusive even before I reached the part of the script on isolation requirements. I don't blame them - it's hard emotionally to be told you are a close contact.

THANK YOU TO OUR
SPONSORS!

BIO
STATISTICS
COLLABORATION
OF AUSTRALIA

sdas

Arbutus | QDA Miner | **Stata** | Stat/Transfer | WordStat



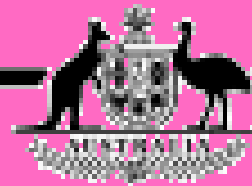
ACEMS

AUSTRALIAN RESEARCH COUNCIL CENTRE OF EXCELLENCE FOR
MATHEMATICAL AND STATISTICAL FRONTIERS



AMSI

AUSTRALIAN
MATHEMATICAL
SCIENCES
INSTITUTE



Australian
Bureau of
Statistics