

Wombat - 2016

Simple tools for complex problems: making molehills out of mountains

Dr Zoe van Havre

Urgency?

Hope. . .

Who am I?

- ▶ PhD in statistics, from QUT & Paris-Dauphine

Who am I?

- ▶ PhD in statistics, from QUT & Paris-Dauphine
- ▶ I live in Brisbane, by way of Canada, New Zealand, and various places in between.

Who am I?

- ▶ PhD in statistics, from QUT & Paris-Dauphine
- ▶ I live in Brisbane, by way of Canada, New Zealand, and various places in between.
- ▶ *Key areas:*

Who am I?

- ▶ PhD in statistics, from QUT & Paris-Dauphine
- ▶ I live in Brisbane, by way of Canada, New Zealand, and various places in between.
- ▶ *Key areas*:
 - ▶ Bayesian statistics

Who am I?

- ▶ PhD in statistics, from QUT & Paris-Dauphine
- ▶ I live in Brisbane, by way of Canada, New Zealand, and various places in between.
- ▶ *Key areas*:
 - ▶ Bayesian statistics
 - ▶ Mixture and hidden Markov models,

Who am I?

- ▶ PhD in statistics, from QUT & Paris-Dauphine
- ▶ I live in Brisbane, by way of Canada, New Zealand, and various places in between.
- ▶ *Key areas:*
 - ▶ Bayesian statistics
 - ▶ Mixture and hidden Markov models,
 - ▶ Bio-statistics/informatics/security,

Who am I?

- ▶ PhD in statistics, from QUT & Paris-Dauphine
- ▶ I live in Brisbane, by way of Canada, New Zealand, and various places in between.
- ▶ *Key areas:*
 - ▶ Bayesian statistics
 - ▶ Mixture and hidden Markov models,
 - ▶ Bio-statistics/informatics/security,
- ▶ *Research interests*

Who am I?

- ▶ PhD in statistics, from QUT & Paris-Dauphine
- ▶ I live in Brisbane, by way of Canada, New Zealand, and various places in between.
- ▶ *Key areas:*
 - ▶ Bayesian statistics
 - ▶ Mixture and hidden Markov models,
 - ▶ Bio-statistics/informatics/security,
- ▶ *Research interests*
 - ▶ data driven, accessible, intuitive tools

Who am I?

- ▶ PhD in statistics, from QUT & Paris-Dauphine
- ▶ I live in Brisbane, by way of Canada, New Zealand, and various places in between.
- ▶ *Key areas:*
 - ▶ Bayesian statistics
 - ▶ Mixture and hidden Markov models,
 - ▶ Bio-statistics/informatics/security,
- ▶ *Research interests*
 - ▶ data driven, accessible, intuitive tools
 - ▶ **making data analysis easier**

What drives me?

The most common question asked since I started to pursue Statistics has been

“Why...?”

- ▶ I have three reasons:

What drives me?

The most common question asked since I started to pursue Statistics has been

“Why...?”

► I have three reasons:

1. A sense of urgency,

What drives me?

The most common question asked since I started to pursue Statistics has been

“Why...?”

► I have three reasons:

1. A sense of urgency,
2. tantalizing hope,

What drives me?

The most common question asked since I started to pursue Statistics has been

“Why...?”

► I have three reasons:

1. A sense of urgency,
2. tantalizing hope,
3. boundless excitement.

Urgency?

Can we keep up?

- ▶ The exponential growth of computing has not slowed down.

Can we keep up?

- ▶ The exponential growth of computing has not slowed down.
- ▶ New types of data and new challenges require new approaches

Can we keep up?

- ▶ The exponential growth of computing has not slowed down.
- ▶ New types of data and new challenges require new approaches
- ▶

Hope...

Not all change is bad

- ▶ Everyone is coming onboard! amazing advances

Not all change is bad

- ▶ Everyone is coming onboard! amazing advances
- ▶ Data-science is a thing now

Not all change is bad

- ▶ Everyone is coming onboard! amazing advances
- ▶ Data-science is a thing now
- ▶ We are standing on a methodological goldmine. . .

Not all change is bad

- ▶ Everyone is coming onboard! amazing advances
- ▶ Data-science is a thing now
- ▶ We are standing on a methodological goldmine. . .
 - ▶ **the traditional way:**

Not all change is bad

- ▶ Everyone is coming onboard! amazing advances
- ▶ Data-science is a thing now
- ▶ We are standing on a methodological goldmine. . .
 - ▶ **the traditional way:**
 - ▶ Develop methods based on large sample theory.

Not all change is bad

- ▶ Everyone is coming onboard! amazing advances
- ▶ Data-science is a thing now
- ▶ We are standing on a methodological goldmine. . .
 - ▶ **the traditional way:**
 - ▶ Develop methods based on large sample theory.
 - ▶ Adapt / make assumptions. to deploy on realistic sample sizes

Not all change is bad

- ▶ Everyone is coming onboard! amazing advances
- ▶ Data-science is a thing now
- ▶ We are standing on a methodological goldmine. . .
 - ▶ **the traditional way:**
 - ▶ Develop methods based on large sample theory.
 - ▶ Adapt / make assumptions. to deploy on realistic sample sizes
 - ▶ **the future?**

Not all change is bad

- ▶ Everyone is coming onboard! amazing advances
- ▶ Data-science is a thing now
- ▶ We are standing on a methodological goldmine. . .
 - ▶ **the traditional way:**
 - ▶ Develop methods based on large sample theory.
 - ▶ Adapt / make assumptions. to deploy on realistic sample sizes
 - ▶ **the future?**
 - ▶ Rework common tools to be closer to underlying theory

Not all change is bad

- ▶ Everyone is coming onboard! amazing advances
- ▶ Data-science is a thing now
- ▶ We are standing on a methodological goldmine. . .
 - ▶ **the traditional way:**
 - ▶ Develop methods based on large sample theory.
 - ▶ Adapt / make assumptions. to deploy on realistic sample sizes
 - ▶ **the future?**
 - ▶ Rework common tools to be closer to underlying theory
 - ▶ This usually means Bayesian, yes. Sorry.

Not all change is bad

- ▶ Everyone is coming onboard! amazing advances
- ▶ Data-science is a thing now
- ▶ We are standing on a methodological goldmine. . .
 - ▶ **the traditional way:**
 - ▶ Develop methods based on large sample theory.
 - ▶ Adapt / make assumptions. to deploy on realistic sample sizes
 - ▶ **the future?**
 - ▶ Rework common tools to be closer to underlying theory
 - ▶ This usually means Bayesian, yes. Sorry.
 - ▶ Asymptotic theory → Methods → Big Data → Theoretically Supported Results

3. Excitement!

Amazing things happen when data analysis combines

- ▶ clear research questions,

3. Excitement!

Amazing things happen when data analysis combines

- ▶ clear research questions,
- ▶ suitable tools, and

3. Excitement!

Amazing things happen when data analysis combines

- ▶ clear research questions,
- ▶ suitable tools, and
- ▶ appropriate data

A short story | overfitted mixtures and Alzheimer's Disease

Key background

What you need to know

- ▶ Alzheimer's Disease (AD) is something we need to address

Key background

What you need to know

- ▶ Alzheimer's Disease (AD) is something we need to address
- ▶ disease development is very slow

Key background

What you need to know

- ▶ Alzheimer's Disease (AD) is something we need to address
- ▶ disease development is very slow
- ▶ no cognitive changes for ≥ 20 years

Key background

What you need to know

- ▶ Alzheimer's Disease (AD) is something we need to address
- ▶ disease development is very slow
- ▶ no cognitive changes for ≥ 20 years
- ▶ once changes evident, extensive physical damage

Key background

What you need to know

- ▶ Alzheimer's Disease (AD) is something we need to address
- ▶ disease development is very slow
- ▶ no cognitive changes for ≥ 20 years
- ▶ once changes evident, extensive physical damage
 - ▶ Amyloid β deposits

Key background

What you need to know

- ▶ Alzheimer's Disease (AD) is something we need to address
- ▶ disease development is very slow
- ▶ no cognitive changes for ≥ 20 years
- ▶ once changes evident, extensive physical damage
 - ▶ Amyloid β deposits
 - ▶ Neuro tangles

Key background

What you need to know

- ▶ Alzheimer's Disease (AD) is something we need to address
- ▶ disease development is very slow
- ▶ no cognitive changes for ≥ 20 years
- ▶ once changes evident, extensive physical damage
 - ▶ Amyloid β deposits
 - ▶ Neuro tangles
 - ▶ Atrophy

Key background

What you need to know

- ▶ Alzheimer's Disease (AD) is something we need to address
- ▶ disease development is very slow
- ▶ no cognitive changes for ≥ 20 years
- ▶ once changes evident, extensive physical damage
 - ▶ Amyloid β deposits
 - ▶ Neuro tangles
 - ▶ Atrophy
- ▶ Tests which assess physical change are \$ \$ \$ and intrusive

How can we help improve early detection

To better tackle AD, we need to be able to treat it earlier.

- ▶ we know little about how AD behaves in its early stage

How can we help improve early detection

To better tackle AD, we need to be able to treat it earlier.

- ▶ we know little about how AD behaves in its early stage
- ▶ could compare known cases to controls,

How can we help improve early detection

To better tackle AD, we need to be able to treat it earlier.

- ▶ we know little about how AD behaves in its early stage
- ▶ could compare known cases to controls,
 - ▶ does not target early stage of AD

How can we help improve early detection

To better tackle AD, we need to be able to treat it earlier.

- ▶ we know little about how AD behaves in its early stage
- ▶ could compare known cases to controls,
 - ▶ does not target early stage of AD
- ▶ **would like to identify individuals likely to be in early stage of AD**

How?

- ▶ large repository of data exists thanks to AIBL study

How?

- ▶ large repository of data exists thanks to AIBL study
- ▶ many data types, potential variables, time points, and sources

How?

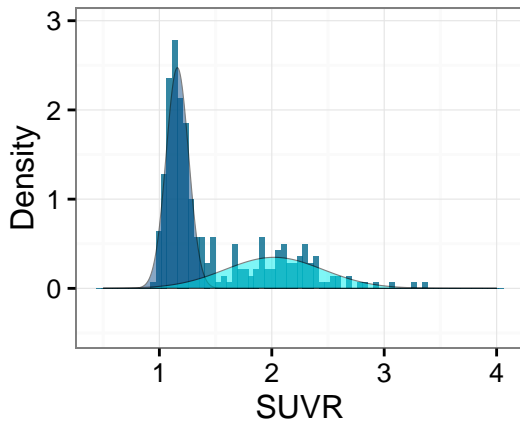
- ▶ large repository of data exists thanks to AIBL study
- ▶ many data types, potential variables, time points, and sources
- ▶ possibilities = *endless* (thousands of potential approaches)

How?

- ▶ large repository of data exists thanks to AIBL study
- ▶ many data types, potential variables, time points, and sources
- ▶ possibilities = *endless* (thousands of potential approaches)
- ▶ What now...?

Ventrolateral.Prefrontal

HC (Blue)



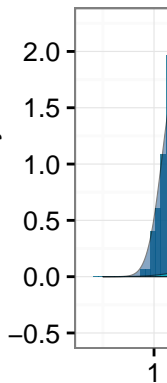
Cluster



HC_k_1

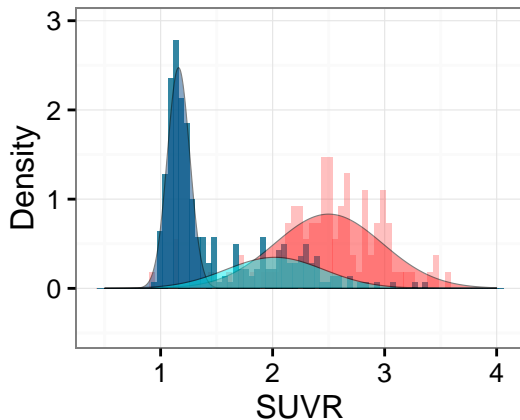
HC_k_2

Density

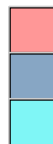


Ventrolateral.Prefrontal

HC (Blue) & AD (Red)

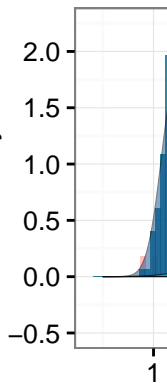


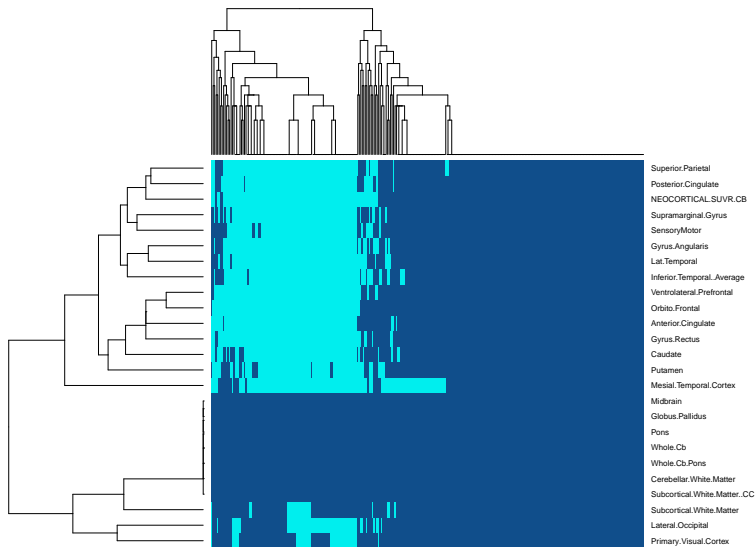
Cluster



AD_k_1
HC_k_1
HC_k_2

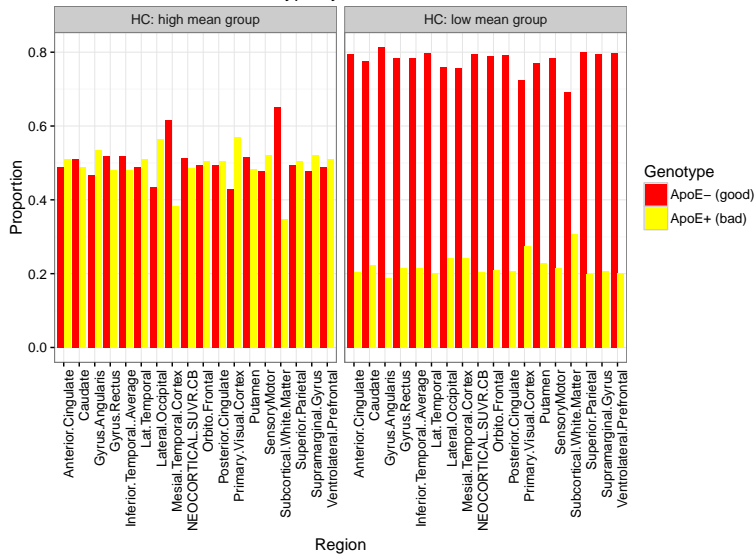
Density



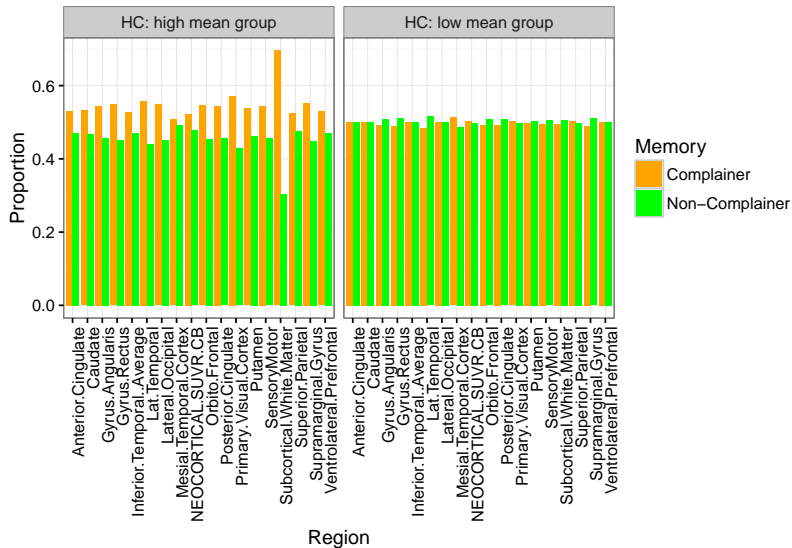


Patients (HC only)

Genotype by Zmix cluster



Memory Status by Zmix cluster



Slide with image