

# Project work

- Choose a data set and make all the steps of Bayesian data analysis workflow listed below
- Project outcome is a R or Python notebook similar to notebooks in (many of these notebooks don't have all the required parts)
  - BDA R demos [https://github.com/avehtari/BDA\\_R\\_demos/tree/master/demos\\_rstan](https://github.com/avehtari/BDA_R_demos/tree/master/demos_rstan)
  - BDA Python demos [https://github.com/avehtari/BDA\\_py\\_demos/tree/master/demos\\_pystan](https://github.com/avehtari/BDA_py_demos/tree/master/demos_pystan)
  - Stan case studies <http://mc-stan.org/users/documentation/case-studies.html>
  - StanCon case studies <http://mc-stan.org/users/documentation/case-studies.html>  
(some of these notebooks are for a bigger projects, but reflect still the basic idea of a notebook presentation)
- The submitted notebooks need to illustrate the knowledge of the Bayesian workflow.

# Project work

- The notebooks have to include
  - Description of the data, and the analysis problem
  - Description of at least two models, for example:
    - non-hierarchical and hierarchical
    - linear and non-linear
    - variable selection with many models
  - Informative or weakly informative priors, and description of the prior choices
  - Stan code
  - How Stan model is run
  - Convergence diagnostics (Rhat, divergences, neff)
  - Posterior predictive checking
  - Model comparison (e.g. with loo)
  - Predictive performance assessment if applicable (e.g. classification accuracy)
  - Sensitivity analysis with respect to prior choices
  - Discussion of problems, and potential improvements
    - It is possible that your model or inference is not perfect, but a better model would require substantial work. Then it's ok that you report the problems found (using the various diagnostics discussed in the course) and describe possible

# Project work

- You can re-use of code and text from existing case studies
  - Just report what did you re-use
  - Acknowledge the original authors
  - Include the original copyright licence
    - CC-BY or CC-BY-NC is common for text  
<https://creativecommons.org/licenses/>
    - BSD-3 is common for code  
<https://opensource.org/licenses/BSD-3-Clause>
  - Don't use improper priors even if some case study has improper priors
- You can use BRMS to create Stan code, but do not limit yourself to BRMS models if changes would make a better model

## Oral presentation

- During evaluation week 50
- Each project needs to be presented in addition to submitting the notebook
- The presentation should be high level but sufficiently detailed information should be readily available to facilitate answering questions from the audience
- Within each session, about four groups will be presenting
- For 1-2 person groups, the presentation should be 10 minutes
- For 3 person groups, the presentation should be 15 minutes
- Afterwards, questions will be asked first by other students and then by two attending TAs for about 5 to 10 minutes
- Grading of the presentation will be done by the two TAs using standardized grading instructions

## Some special topics

- Improve R demos
- Dynamic HMC demo in R or Python

## Some ideas for data sets

- How do People Type on Mobile Devices? Observations from a Study with 37,000 Volunteers  
<https://userinterfaces.aalto.fi/typing37k/>
- Laptop multitasking hinders classroom learning for both users and nearby peers <http://www.sciencedirect.com/science/article/pii/S0360131512002254>
- Arctic sea ice shrinking  
<https://www.nytimes.com/interactive/2017/09/22/climate/arctic-sea-ice-shrinking-trend-watch.html>
- Finnish weather statistics  
<https://en.ilmatieteenlaitos.fi/statistics-from-1961-onwards>
- R datasets <https://vincentarelbundock.github.io/Rdatasets/datasets.html>
- Vanderbilt Biostatistics  
<http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets>
- Probably better to \*not\* have a data set
  - with number of observation in millions
  - machine vision task

# Schedule

- Register project group and topic by 4th November
  - 2 person group preferred, 3 and 1 person groups allowed
  - 3 person groups are expected to choose more difficult projects
- During the week starting 4th November, start working on the project and if necessary talk with TAs (no new assignment on that week)
- Deadline end of week 49, 8 December
- Oral presentations during the evaluation week (week 50)