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)
    - 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)
    - More examples in
      https://avehtari.github.io/BDA_course_Aalto/project.html
- Max 20 pages (less is fine, the notebook format is loose, and code, output and figures take space, too)
- The submitted notebooks need to illustrate the knowledge of the Bayesian workflow.

Project work

- The notebooks have to include, see
  https://avehtari.github.io/BDA_course_Aalto/project.html
  - 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, rstanarm, or brms code
  - How Stan inference is run
  - Convergence diagnostics (Rhat, divergences, neff)
  - Posterior predictive checking
  - Model comparison (e.g. with loo)
  - Predictive performance assessment if applicable (bonus points)
  - 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

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
  - Small amount of verbatim quotes of the text allowed when appropriately marked, but otherwise use your own words
  - 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 rstanarm and brms, but do not limit yourself to rstanarm or brms models if changes would make a better model

Oral presentation

- Each project needs to be presented in addition to submitting the notebook
- Peergrading of the notebooks happen before the presentations, so you get feedback and you can improve for the presentation
- 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

Some ideas for data sets

- 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

- Probaly better to **not** have a data set
  - with number of observation in millions
  - machine vision task

Schedule

- Register project group and topic
  - 3 person groups preferred, 1–2 person groups allowed
  - 3 person groups are expected to choose a slightly more difficult projects
  - 3 person groups are highly recommended over 1–2 person groups. 3 person groups have priority when reserving presentation slots
- I will provide further advice on presentations later