Bayesian Hierarchical Model for Maize Yield

Alice Huang

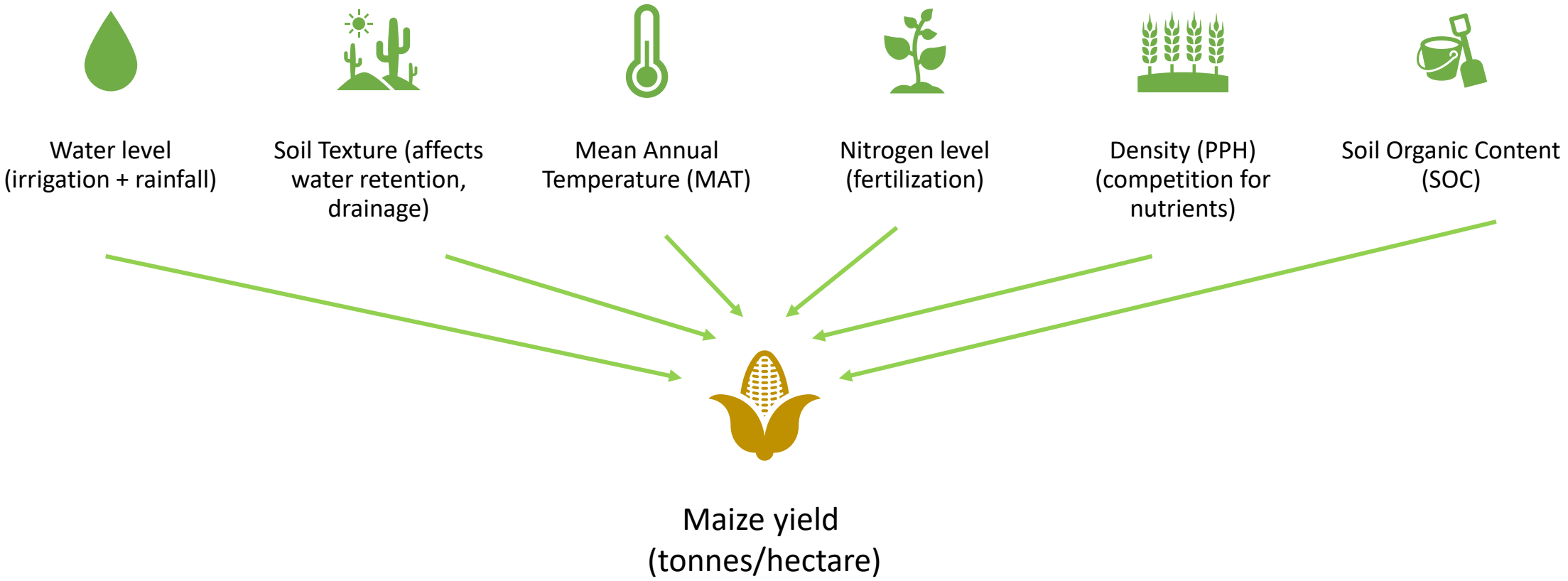
Introduction

- Maize (a.k.a corn in North America) is one of the most commonly consumed crops in the world
- Used to make animal feed, cornstarch, corn syrup, and more
- We are interested in factors that influence its yield
- Implications for business profitability, food shortages

Data

- Downloaded from https://data.mendeley.com/datasets/7m3pjrz52x
- Collection of data from many different studies on maize yield
- Featured in paper by Li et al (2019)
  - A global synthesis of the effect of water and nitrogen input on maize (*Zea mays*) yield, water productivity and nitrogen use efficiency
- In this project I focus on maize yield data from China
  - Help control for correlation due to country's agriculture laws & policies, farming techniques, technology
  - More data available for different soil types
  - Data was more complete
  - One of the biggest producers of maize

Potential Covariates Based on Previous Literature



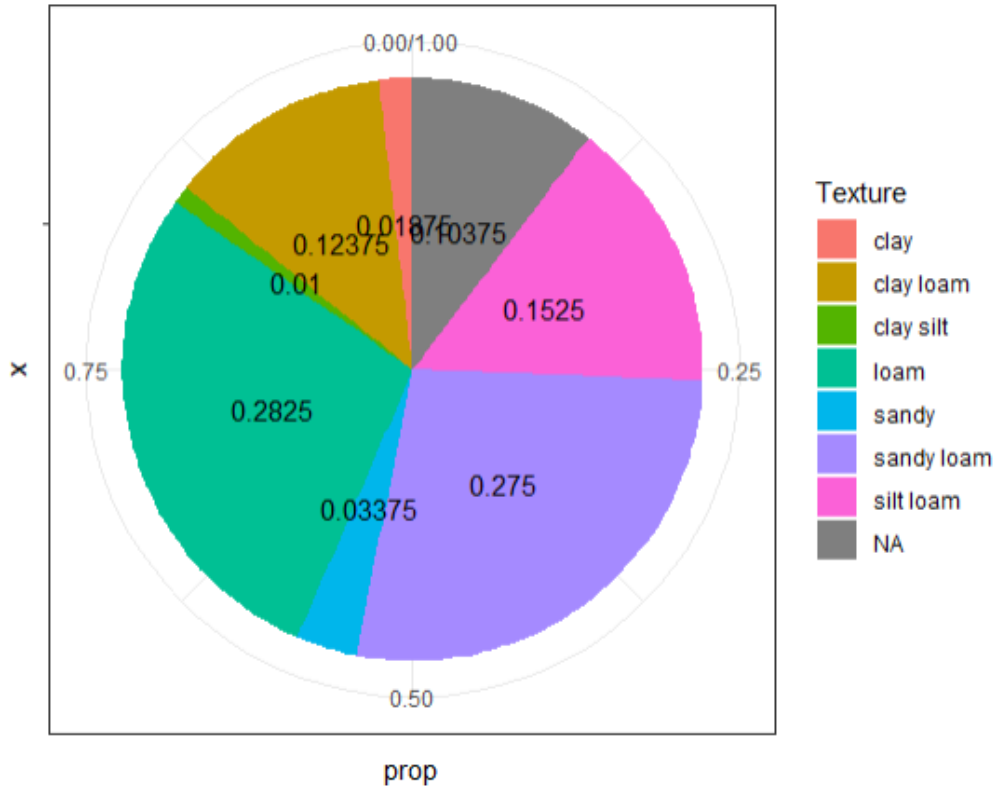
Research Question

- How should we expect maize yield to change given changes in soil type, water level, nitrogen level, soil organic content, mean annual temperature and plant density?

Summary Statistics

Aggregated 800 observations from 52 studies on maize yield in China, 39 locations

Yield	MAT	PPH	SOC	Water	Nitrogen
Min. : 1.000	Min. : 0.500	Min. : 40000	Min. : 4.270	Min. : 144.8	Min. : 0.0
1st Qu.: 6.497	1st Qu.: 7.000	1st Qu.: 56500	1st Qu.: 6.970	1st Qu.: 368.1	1st Qu.: 80.0
Median : 8.585	Median : 9.700	Median : 66700	Median : 7.660	Median : 442.0	Median :180.0
Mean : 8.777	Mean : 9.724	Mean : 68006	Mean : 8.687	Mean : 476.4	Mean :189.9
3rd Qu.:10.600	3rd Qu.:12.900	3rd Qu.: 72000	3rd Qu.: 9.110	3rd Qu.: 534.8	3rd Qu.:280.0
Max. :19.200	Max. :21.700	Max. :120000	Max. :21.810	Max. :1112.5	Max. :560.0
NA	NA	NA's :19	NA's :168	NA	NA

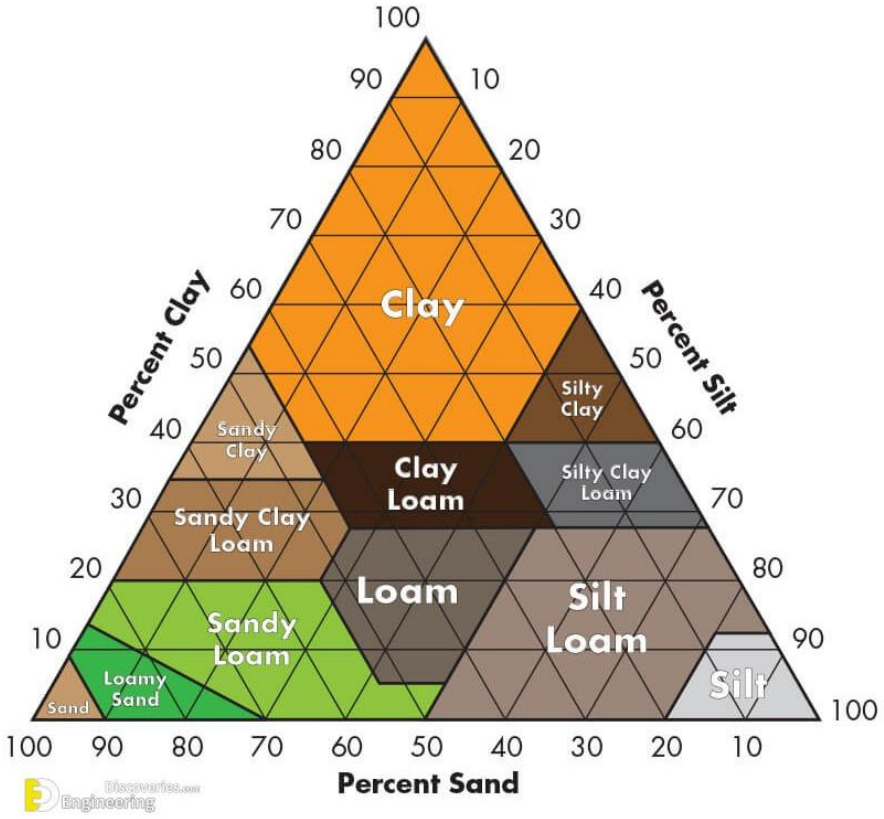


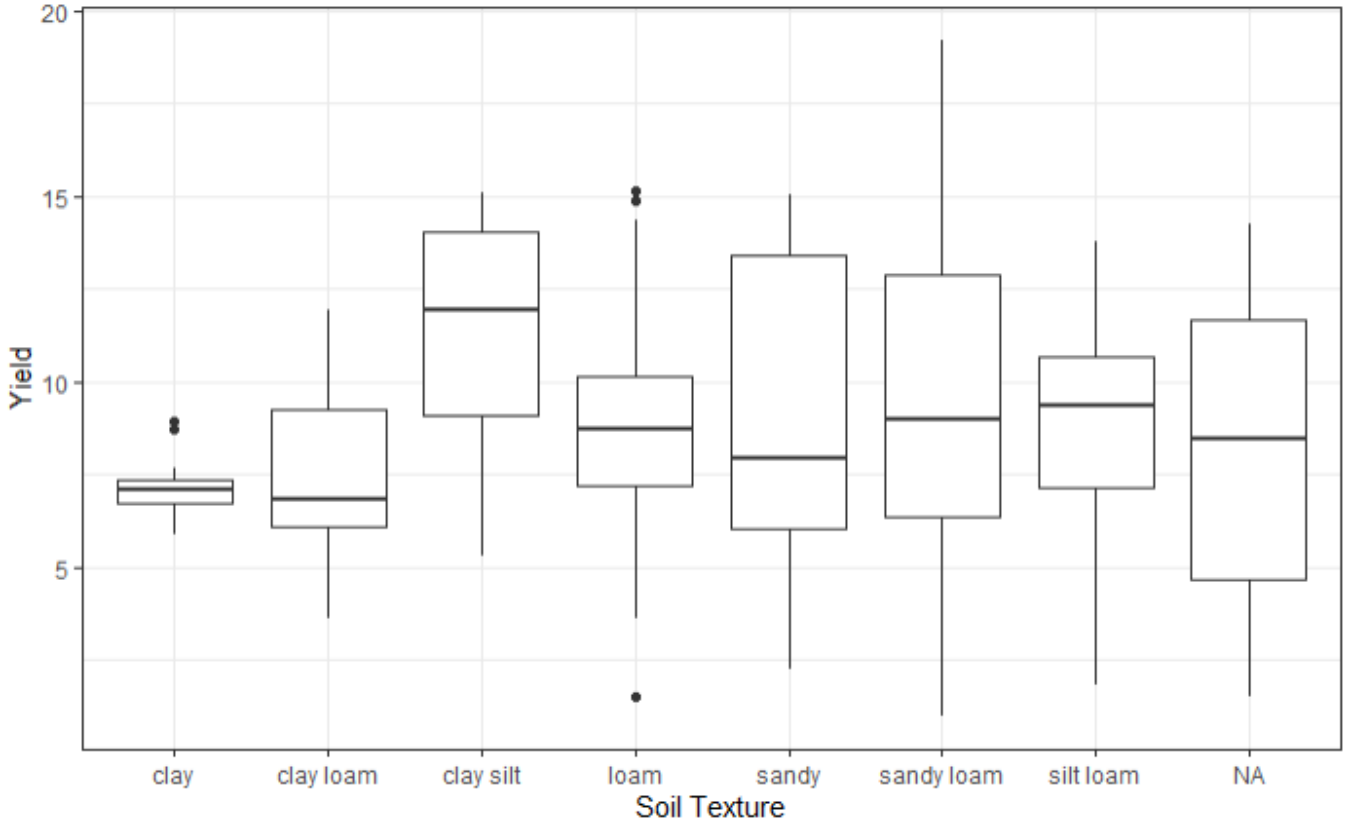
Missing Data

- 80 rows missing Soil Texture type
- Will be a problem as I will try to build a Bayesian hierarchical model with soil being a level in the hierarchy
- May not be able to use this data
- Some soil types missing lots of data on Plants Per Hectare (PPH), don’t have a lot of observations to begin with
- PPH seems to have been kept constant in other soil groups.
- May not be able to use it in model

Soil Type and Maize Yield

- Some types of soil hold onto water better than others
- Important since maize has shallow roots
- Particle size, Water retention: sand < silt < clay
- Soil texture will be recoded into integer factor variable based on expected water retention computed from particle size and percentage of sand, silt, clay

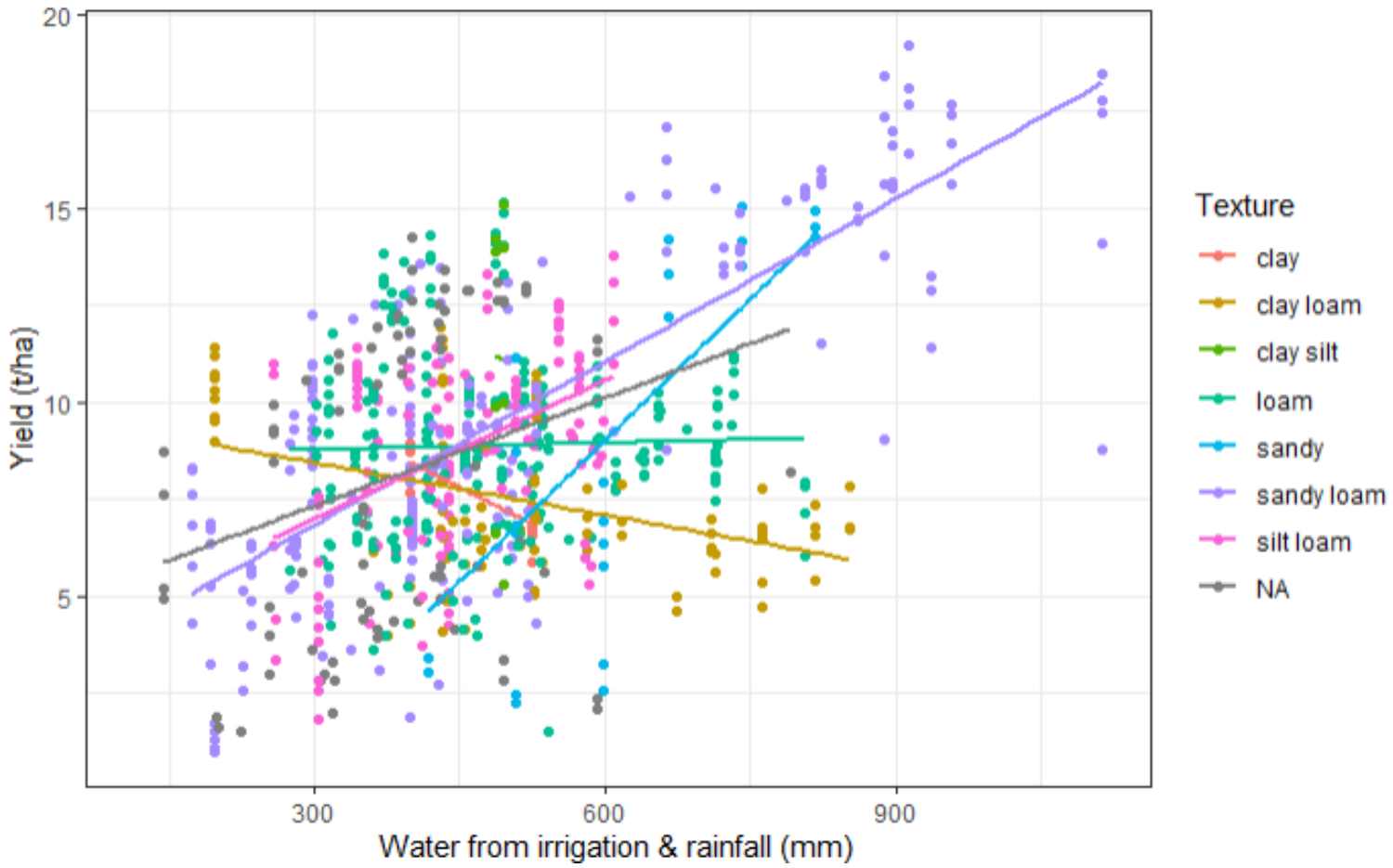




Water and Maize Yield

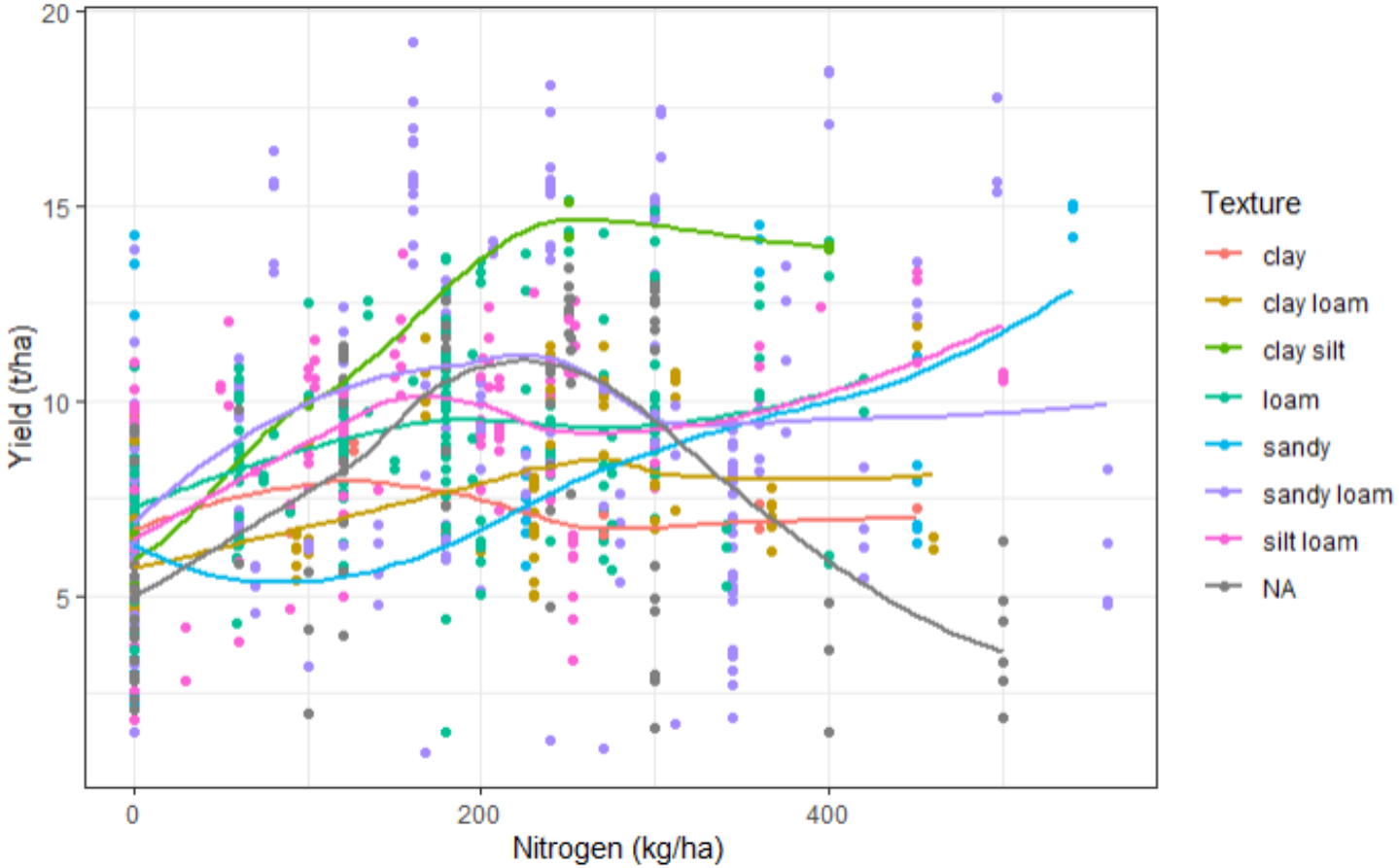
For finer soil types like **clay loam**, increasing water seemed to either have little effect or slightly negative effect on yield

For coarser soil types like **sandy loam**, (worse at retaining water) increasing water seemed to increase yield



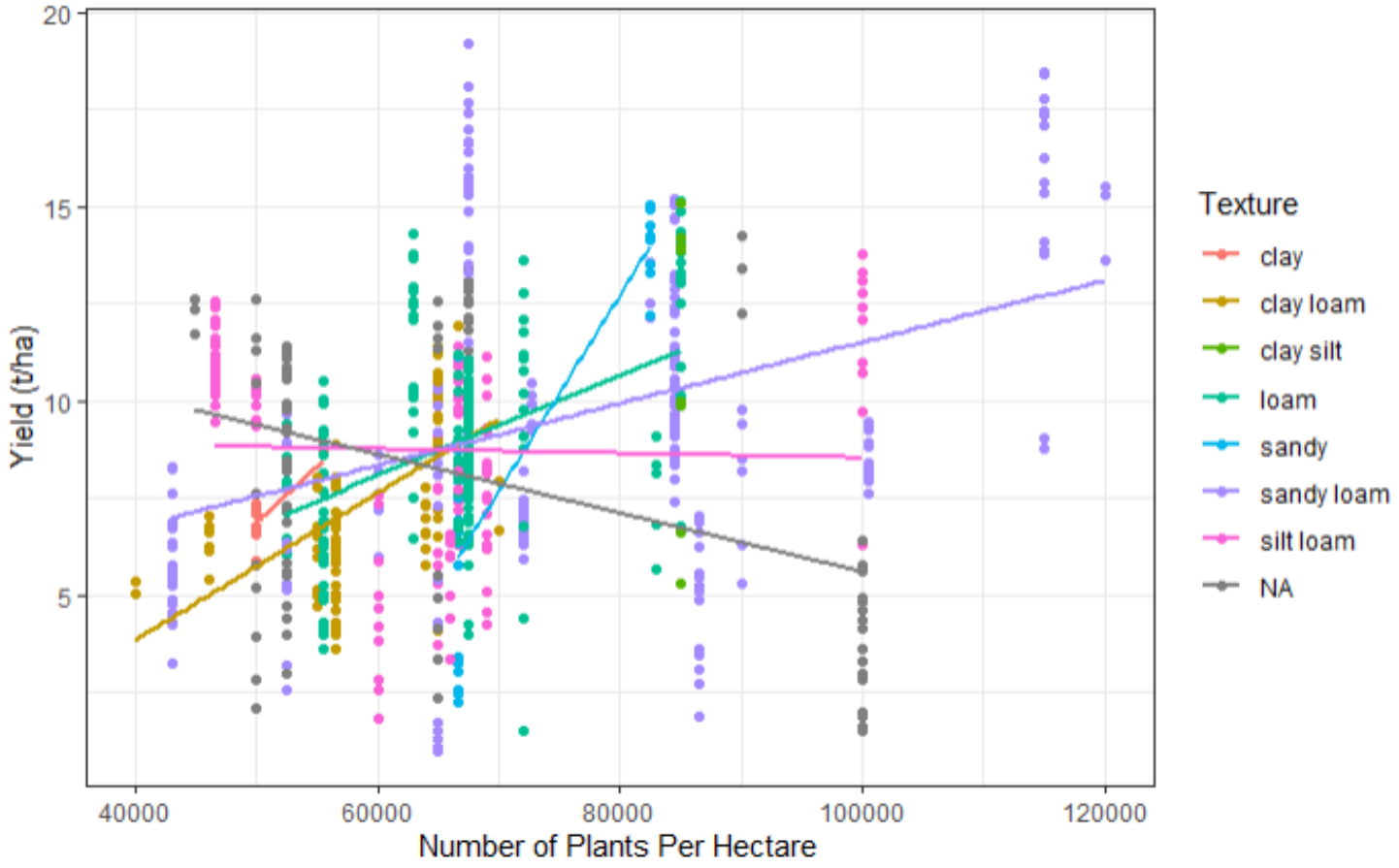
Nitrogen and Maize Yield

- For most soil types, as nitrogen increases, yield either doesn't change much, or increases slightly but plateaus after
- Exception: **sandy** soil, coarsest type which does not hold onto water/nutrients well



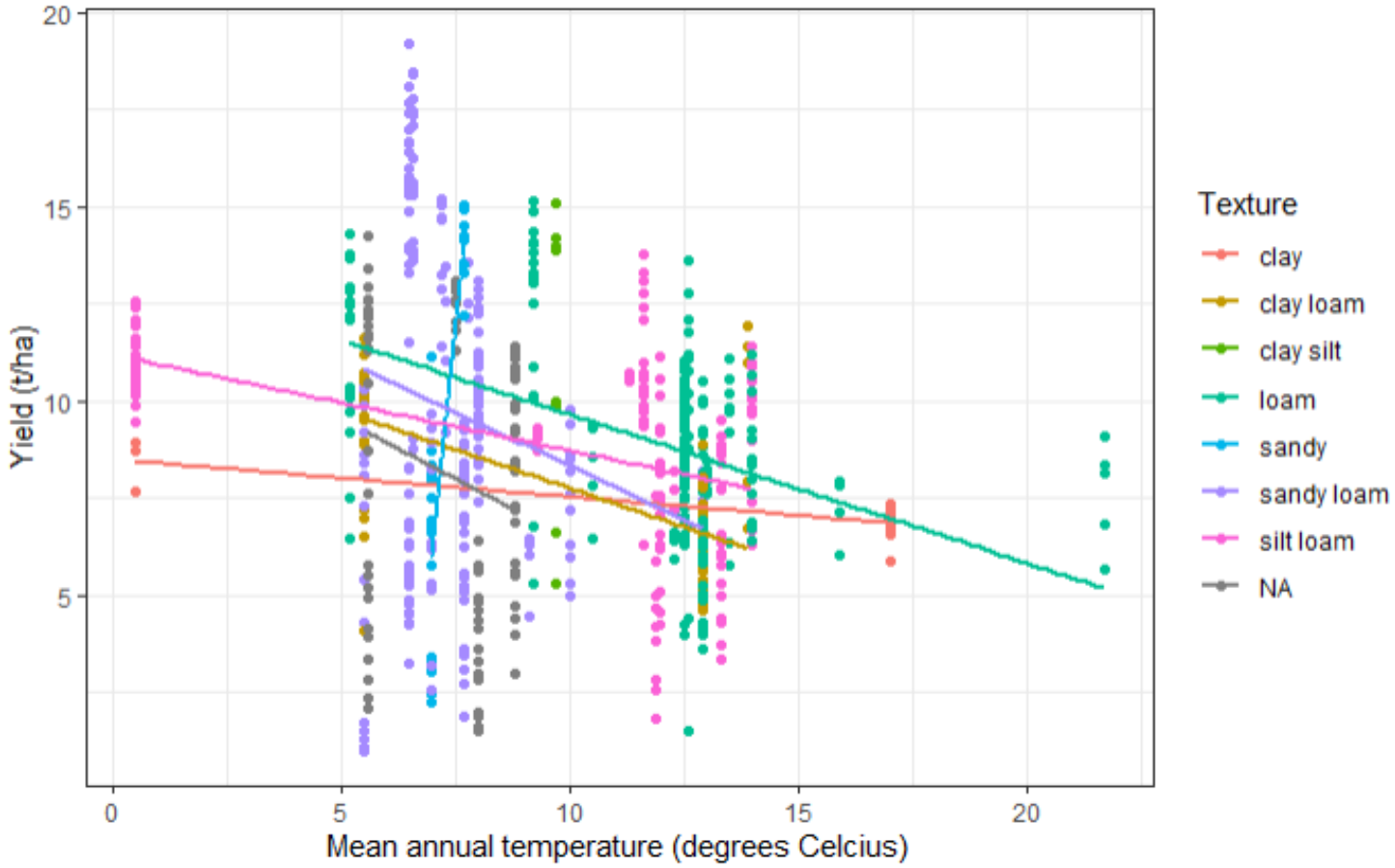
Plant Density and Yield

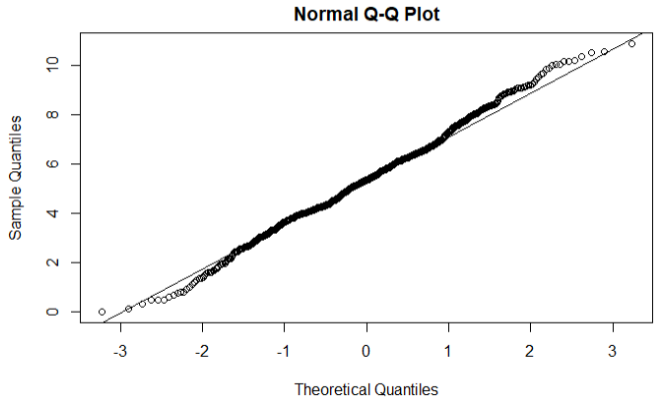
- For most soil types except silt loam, increasing plant density seemed to increase yield
- Perhaps farmers already knew how dense to plant their seeds before plants start competing with each other

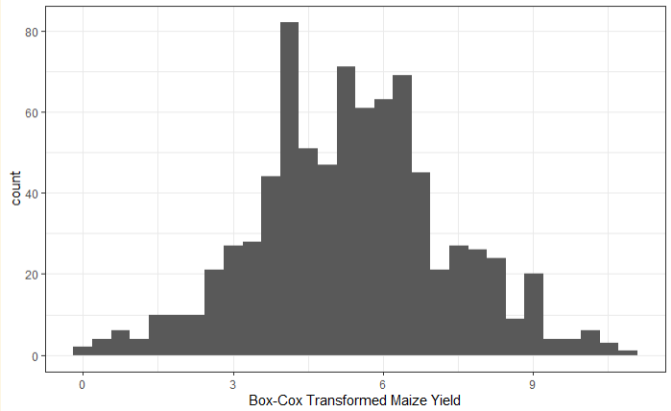


Temperature and Yield

- Increasing temperature seemed to have a negative effect on yield
- Perhaps warmer climates have longer growing seasons -> soil depleted of nutrients
- Soil type and temperatures correlated due to geographic region








```
        Shapiro-Wilk normality test

data:  box(datamaizechina$Yield, bestlambda)
W = 0.99636, p-value = 0.06153
```

Proposed Likelihood for Maize Yield

- Box-cox transformation on maize yield to transform it into a normal variable (simpler for modelling)
- $y \rightarrow \frac{y^{\lambda}-1}{\lambda}$, $\lambda \approx 0.707$
- Normal likelihood for transformed maize yield

Proposed Bayesian Hierarchical Model

$$y_j|\alpha_{j[i]}, \beta_{j[i]}, \gamma_{j[i]}, \sigma_y^2 \sim N(\alpha_{j[i]} + \beta_{j[i]} W_i + \gamma_{j[i]} N_i, \sigma_y^2)$$

y_j is transformed maize yield (t/ha), W_i is water (mm), N_i is nitrogen (kg/ha)

$$\sigma_y^2 \sim N^+(0,1)$$
$$\alpha_j \sim N(\mu_\alpha, \sigma_\alpha^2)$$
$$\mu_\alpha \sim N(0,1)$$
$$\sigma_\alpha^2 \sim N^+(0,1)$$
$$\beta_j \sim N(\mu_\beta, \sigma_\beta^2)$$
$$\mu_\beta \sim N(0,1)$$
$$\sigma_\beta^2 \sim N^+(0,1)$$
$$\gamma_j \sim N(\mu_\gamma, \sigma_\gamma^2)$$
$$\mu_\gamma \sim N(0,1)$$
$$\sigma_\gamma^2 \sim N^+(0,1)$$

- Variables will be centered
- No correlation between water, nitrogen, density
- Temperature correlated with soil
- No association between SOC and yield, other variables
- Stratify on soil type (ordered by particle size)
- Within each soil type, consider effects of water, nitrogen, (and maybe plant density if dealing with missing data is manageable)