TEKNOLOGISK INSTITUT

Notat: Spørgeskemaundersøgelse om Lupinmarker

23-12-2022
Sandra Simone Schluttenhofer, Konsulent
sasr@teknologisk.dk, +45 72 20 22 34

Objective

- Based on answers from a questionnaire survey about the cultivation of lupin in Denmark in 2021 and 2022, we investigate which variables have an effect on the lupin yield and quantify the effect of these variables.

Overall conclusions

- The analysis shows that the following variables have a significant effect on the yield of lupin.
    - **Sådybde**: Sowing Lupin at a medium sådybde (3.5cm – 4.0cm) gives a significantly higher yield than sowing "low" or "high".
    - **Blandet_art:** The kind of crop that lupin is blend with has a significant effect on the yield. Vårhvede/Hvede has a higher yield than pure lupin or a mix with Vårtriticale.
    - **Sådato:** Early sowing gives a significantly higher yield than late sowing.
    - **Husdyrgødning:** Husdyrgødning seems to have a significant positive effect on the yield. There are, however, very few observations with husdyrgødning.

  The effect of all other variables was not significant in our analysis.

Remarks

- The analysis is based on an observational study, which is usually considered to be less generalizable than a planned trial.
- There are too few data points to satisfactory model interactions between all variables.

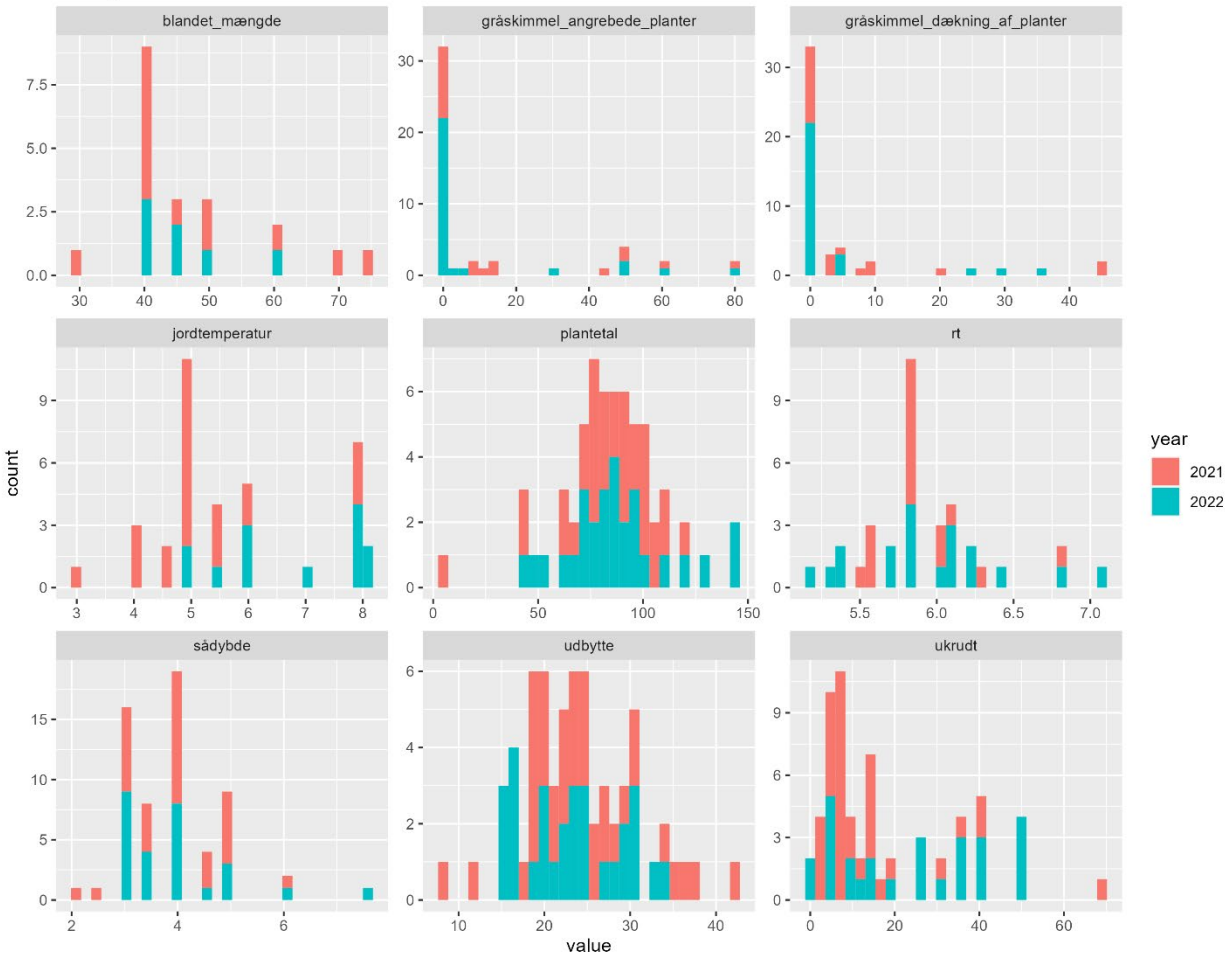
Data

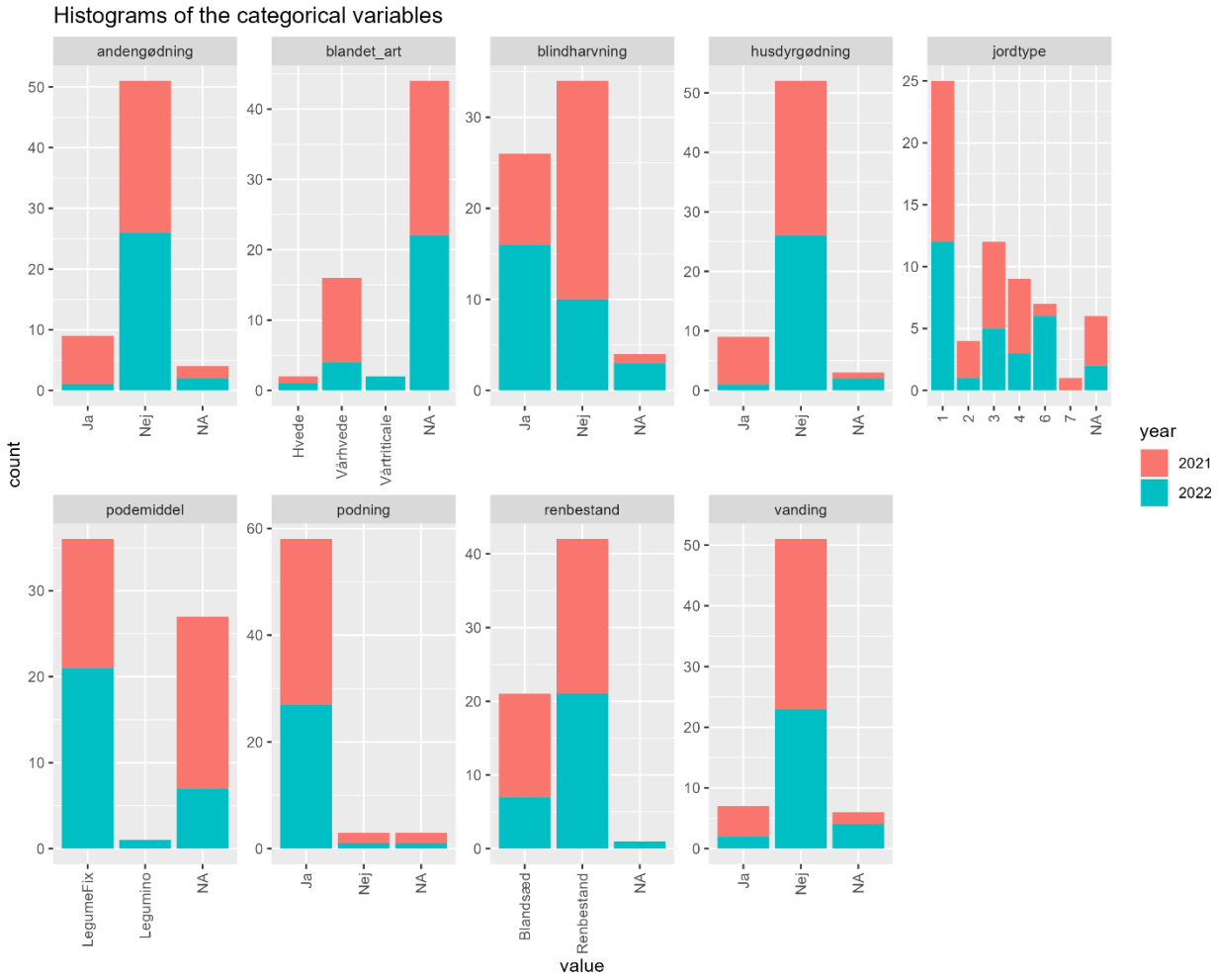
- The data consists of the answers of a questionnaire survey. Each row in the dataset represents a field. There are 36 rows with data from the year 2021 and 30 rows with data from 2022. For both years there exists a field with no information about the lupin yield, which is excluded from the analysis. There are many variables in the data set, in our analysis we include the following (we state the variable name and the corresponding survey question in parentheses).
    - response variables (outcome variables of interest)
        - udbytte ("Udbytte (hkg pr. ha)")
        - plantetal ("Plantebestand (lupin) planter pr. m2")
    - explanatory variables
        - jordtype ("Jordtype på arealerne med lupin")
        - rt ("RT")

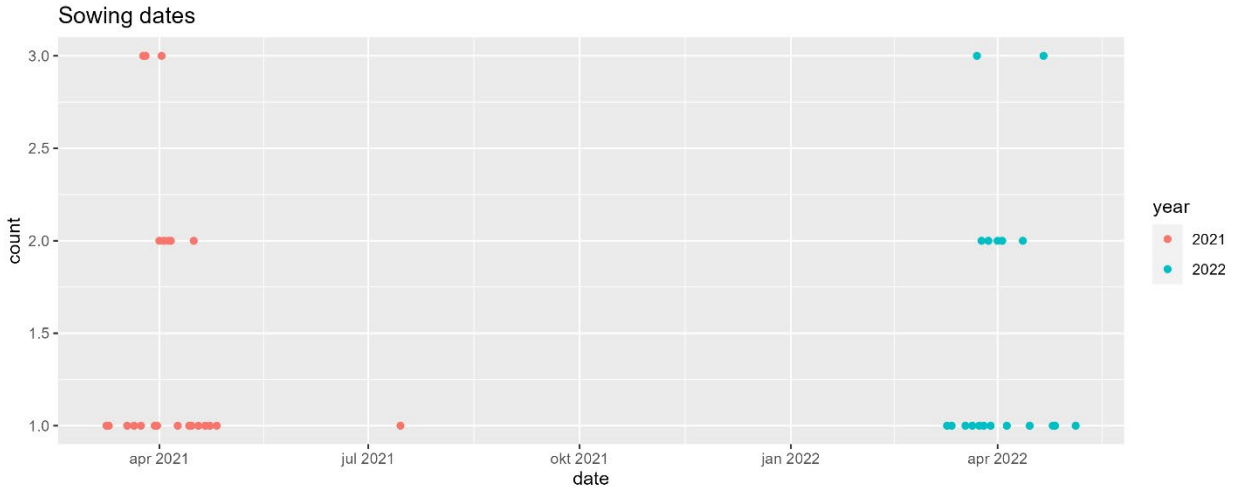
- forfrugt ("Forfrugt til lupin - grupperet")
- forforfrugt ("Angiv forforfrugten til lupin")
- husdyrgødning ("Er der anvendt husdyrgødning?")
- andengødning ("Er der anvendt anden gødning?")
- renbestand ("Har du lupin i renbestand eller blandsæd?")
- sort ("Sortsnavn – Lupin")
- podning ("Er lupin udsæden podet?")
- podemiddel ("Podemiddel anvendt - oprenset")
- blandet_art ("Hvilken anden planteart er lupinen blandet med?")
- blandet_mængde ("Hvilken udsædsmængde er anvendt af denne anden planteart?")
- sådato ("Dato for såning (dd.mm.yyyy)")
- sådybde ("Sådybde (cm)")
- jordtemperatur ("Jordtemperatur ved såning")
- blindharvning ("Er marken blindharvet?")
- vanding ("Er markerne med lupin blevet vandet?")
- ukrudt ("Ukrudt (% dækning af jorden)")
- gråskimmel-angrebede planter ("Gråskimmel (% angrebne planter)")
- gråskimmel-dækning af planter ("Gråskimmel (% dækning af planter)")

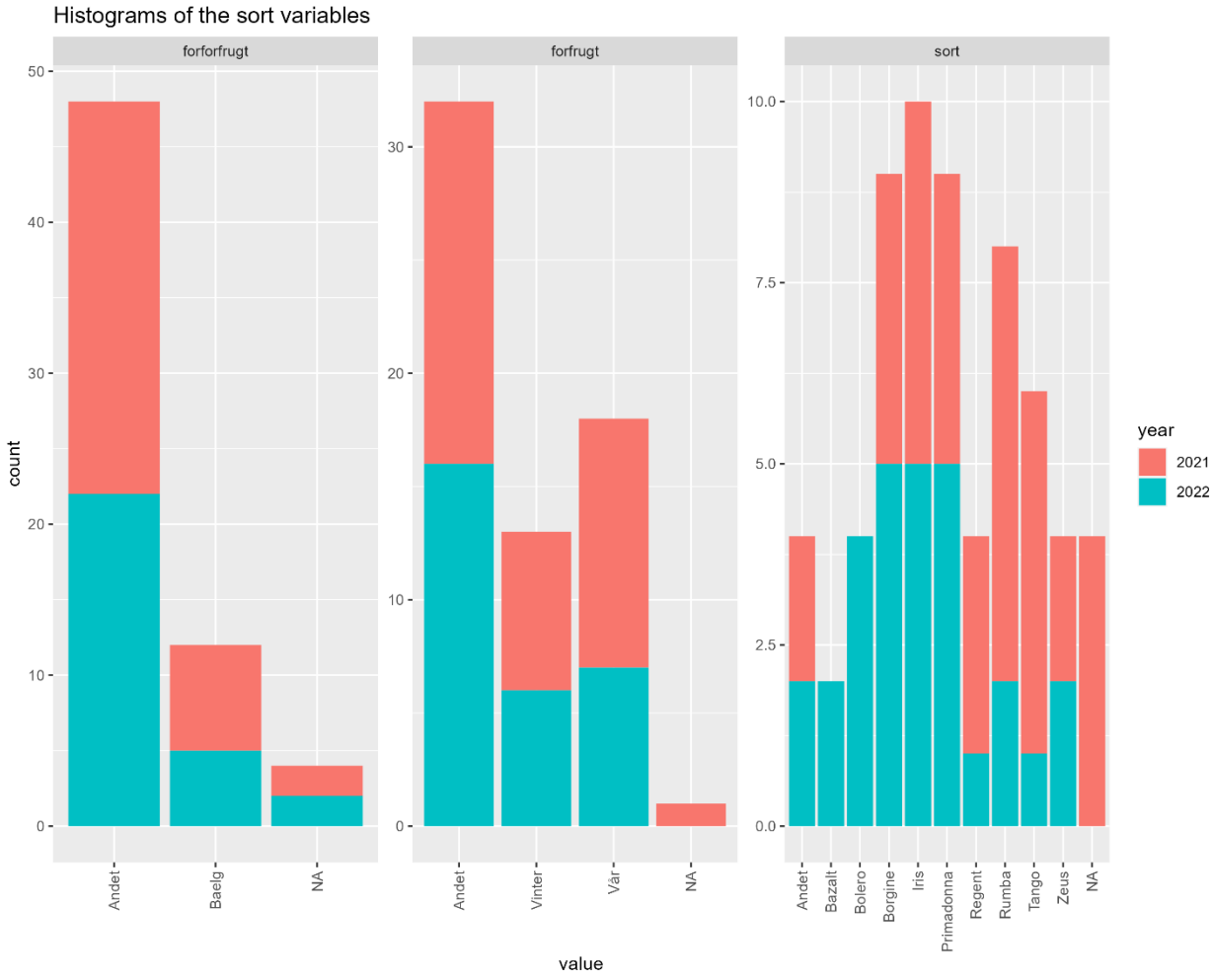
As a first overview over the data we plot histograms for each variable. The colors indicate the year, "NA" stands for "not available/missing data".

Histograms of the numeric variables




Histograms of the categorical variables
andengødning
blandet_art
blindharvning
husdyrgødning
jordtype
podemiddel
podning
renbestand
vanding
count
value
Ja
Nej
NA
Hvede
Vårhvede
Vårtriticale
LegumeFix
Legumino
Blandsæd
Renbestand
1
2
3
4
6
7
year
2021
2022

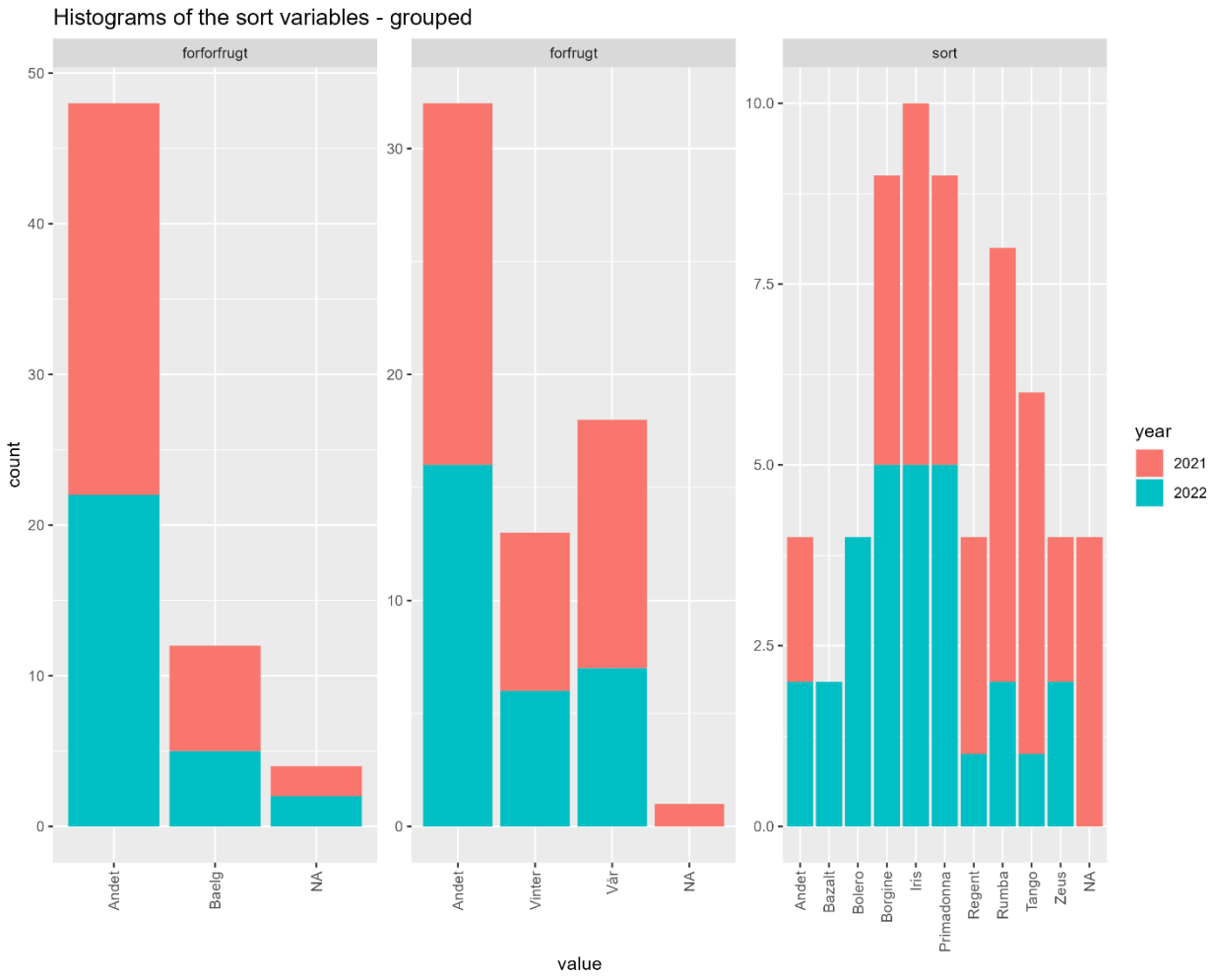

Sowing dates
count
date
apr 2021
jul 2021
okt 2021
jan 2022
apr 2022
year
2021
2022



To make the analysis feasible we have grouped answers for the following variables:

- forfrugt is grouped in "Vintersæd", "Vårsæd" and "Andet"
- forforfrugt is grouped in "Bælgfrugt" and "Andet"
- podning is grouped in "LegumeFix", "Legumino", "Andet" and "Ingen"
- sådato is grouped in "early" (before the end of March) and "late" (after the beginning of April)
- sådybde is grouped in "low" (< 3.5 cm), "medium" (3.5cm – 4.0cm) and "high" (> 4.0cm)
- blandet_mængde is grouped in "low" (<= 40 %), "high" (>40 %) and "none" (no blending)

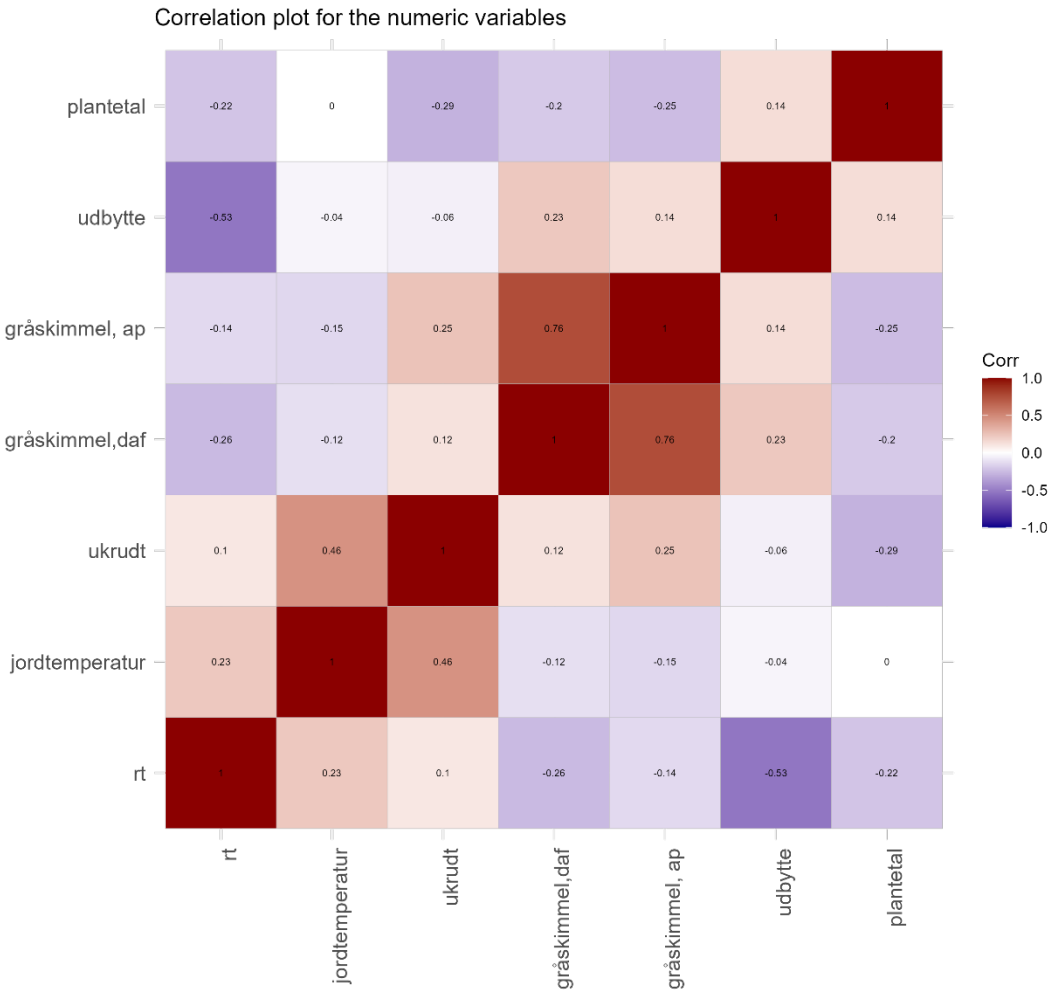
We plot histograms for the grouped variables.



Data exploration

To examine the data we create a **correlation plot** for the numeric variables. The three main observations from the correlation plot are:

- There is a strong positive correlation between "gråskimmel_angrebede_planter" and "gråskimmel_dækning_af_planter". A higher number of infested plants correlates with a higher covering with grey mold.
- There is a positive correlation between "ukrudt" and "jordtemperatur". A higher soil temperature at the sowing date correlates with a higher amount of weed.
- There is a negative correlation between "rt" and "udbytte". A higher reaktionstal correlates with a lower yield.



Data analysis

Part 1: As a first analysis of the data we estimate the effect of each explanatory variable on the response variable "udbytte" in a linear model.

- Tables 1.a – 1.o present the estimated marginal means for the levels of each categorical varialbe, that is the estimated mean value of the response "udbytte" if we use only the categorical variable as an explanatory variable. For the majority of the variables we do not see a significant effect on the yield. The categorical variables with a significant effect are the following:
  - **Renbestand**: There is a significantly higher yield for blandsæd than for renbestand. (Table 1.g)
  - **Blandet_art**: The kind of crop that lupin is blend with has a significant effect on the yield. Vårhvede/Hvede has a higher yield than pure lupin or a mix with Vårtriticale (Table 1.j)
  - **Blandet_mængde**: Blending with a larger amount of another crop has a higher yield. (Tabel 1.k)

- **Sådybde**: Sowing Lupin at a medium sådybde (3.5cm – 4.0cm) gives a significantly higher yield than sowing "low" or "high". (Table 1.l)
- **Sådato**: Early sowing gives a significantly higher yield than late sowing. (Table 1.o)

Tabel 1 a: Effect of year

year	udbytte	Lower CL	Upper CL	Signifikans Gruppe
2021	23.18	20.793	25.567	a
2022	24.412	22.239	26.585	a

Tabel 1 b: Effect of jordtype

jordtype	udbytte	Lower CL	Upper CL	Signifikans Gruppe
7	8	-4.178	20.178	a
6	17.754	13.151	22.357	a
3	23.263	19.748	26.779	a
1	25.196	22.76	27.631	a
2	25.75	19.661	31.839	a
4	25.867	21.807	29.926	a

Tabel 1 c: Effect of forfrugt

forfrugt	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Vinter	21.943	18.395	25.491	a
Vår	22.744	19.729	25.76	a
Andet	25.155	22.894	27.417	a

Tabel 1 d: Effect of forforfrugt

forforfrugt	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Bælg	23.24	19.449	27.031	a
Andet	23.995	22.099	25.89	a

Tabel 1 e: Effect of husdyrgødning

husdyrgødning	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Nej	23.106	21.329	24.884	a
Ja	27.011	22.739	31.283	a

Tabel 1 f: Effect of andengødning

andengødning	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Nej	23.217	21.436	24.998	a
Ja	25.951	21.712	30.19	a

Tabel 1 g: Effect of renbestand

renbestand	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Renbestand	22.848	20.919	24.777	a
Blandsæd	26.19	23.463	28.918	b

Tabel 1 h: Effect of sort

sort	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Bazalt	15.75	7.048	24.452	a
Bolero	20.225	14.072	26.378	a
Zeus	22.25	16.097	28.403	a
Tango	22.667	17.643	27.69	a
Primadonna	23.278	19.176	27.38	a
Regent	23.39	17.237	29.543	a
Andet	24.25	18.097	30.403	a
Rumba	24.925	20.574	29.276	a
Iris	25.504	21.613	29.395	a
Borgine	28.303	24.201	32.405	a

Tabel 1 i: Effect of podemiddel

podemiddel	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Ingen	21.333	13.863	28.804	a
Legumino	21.53	8.591	34.469	a
LegumeFix	23.452	21.296	25.609	a
Andet	24.491	21.668	27.315	a

Tabel 1 j: Effect of blandet_art

blandet_art	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Vårtriticale	15	6.672	23.328	a
Ingen	22.628	20.852	24.403	a
Vårhvede	27.813	24.868	30.757	b
Hvede	28	19.672	36.328	ab

Tabel 1 k: Effect of blandet_mængde

blandet_mængde	udbytte	Lower CL	Upper CL	Signifikans Gruppe
None	22.628	20.787	24.469	a
Low	24.3	20.438	28.162	ab
High	28.8	24.938	32.662	b

Tabel 1 l: Effect of sådybde

sådybde	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Low	20.521	18.373	22.668	a
High	23.499	20.762	26.237	a
Medium	28.163	25.651	30.675	b

Tabel 1 m: Effect of blindharvning

Blindharvning	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Ja	23.083	20.526	25.64	a
Nej	24.367	22.131	26.603	a

Tabel 1 n: Effect of vanding

Vanding	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Nej	23.101	21.3	24.901	a
Ja	27.429	22.568	32.289	a

Tabel 1 o: Effect of sådato

Sådato	udbytte	Lower CL	Upper CL	Signifikans Gruppe
late	21.394	19.463	23.325	a
early	26.883	24.685	29.082	b

- Tables 2.a-2.e present the estimated linear relationship between the yield and the numeric variables. Denote the "udbytte" by Y and the numeric variable by X. We fit the following linear model

$$Y \sim a + bX,$$

where a is the intercept and b is the slope. We present the estimates for the intercept a and the slope b with their corresponding confidence intervals and p values. Most numeric variables do not have a significant (non-zero) slope/effect on the yield (indicated by p-values below 0.05). The only variable for which we find a significant slope is

  - **Reaktionstal (rt).** Fitting a linear model to the data, we find a significant negative slope, i.e. a higher rt value gives a lower yield.

    Remark: The measured rt values are between 5.2 and 7.1. The fitted linear model gives an intercept, which corresponds to a rt value of 0. There is most likely not a linear relationship between rt and yield over all values of rt. We plot the reaktionstal against the yield for visualization.

Tabel 2 a: Effect of jordtemperatur

	estimate	Lower CL	Upper CL	p.value
Intercept	25.033	14.948	35.118	0
slope	-0.208	-1.885	1.469	0.803

Tabel 2 b: Effect of ukrudt

	estimate	Lower CL	Upper CL	p.value
Intercept	24.437	21.956	26.918	0
slope	-0.024	-0.128	0.079	0.64

Tabel 2 c: Effect of rt

	estimate	Lower CL	Upper CL	p.value
Intercept	68.22	43.11	93.331	0
slope	-7.417	-11.653	-3.181	0.001

Tabel 2 d: Effect of plantetal

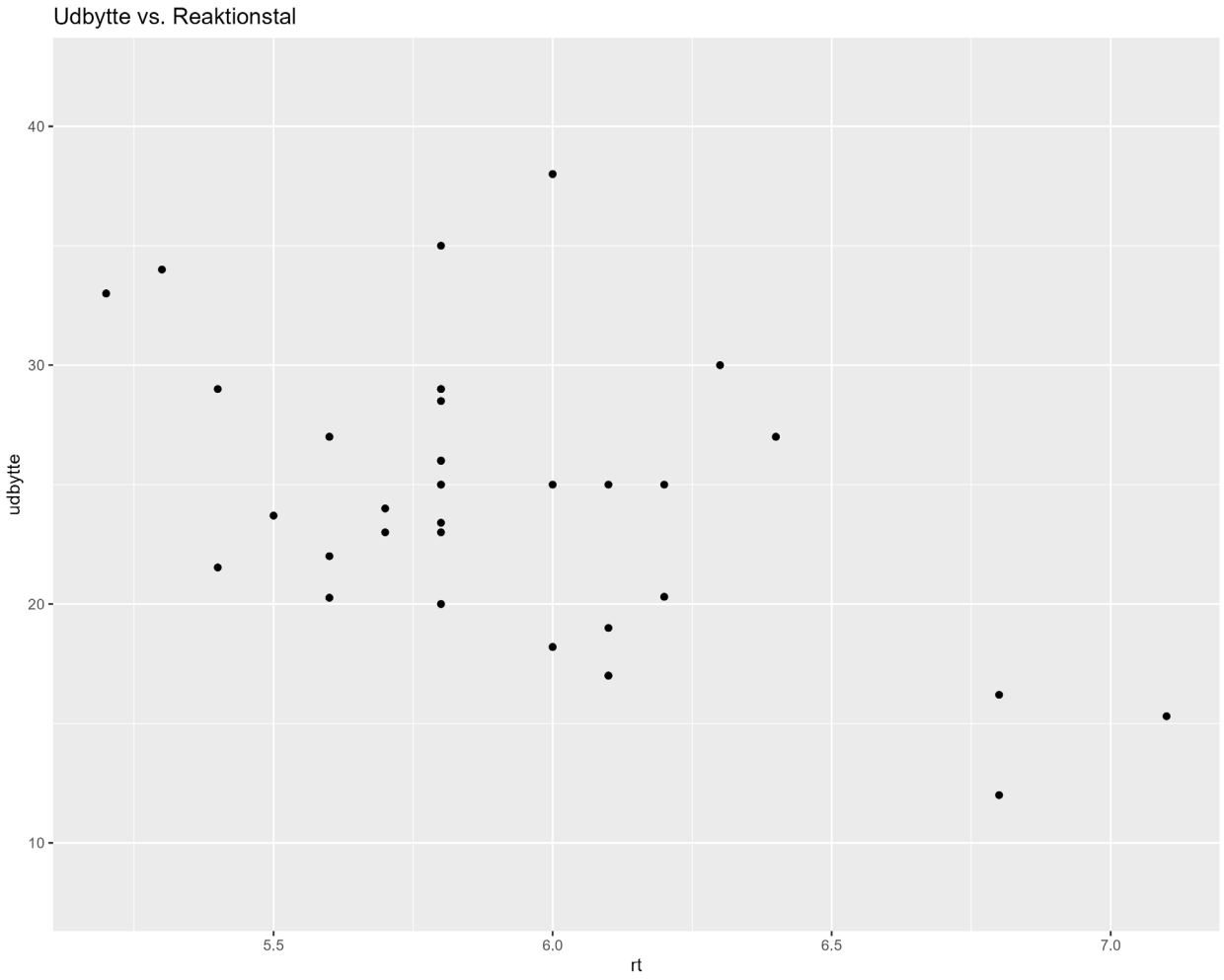
	estimate	Lower CL	Upper CL	p.value
Intercept	20.818	14.591	27.046	0
slope	0.038	-0.032	0.109	0.282

Tabel 2 e: Effect of gråskimmel_angrebede_planter

	estimate	Lower CL	Upper CL	p.value
Intercept	23.183	14.591	21.451	0
slope	0.07	-0.032	-0.006	0.07

Tabel 2 e: Effect of gråskimmel_dækning_af_planter

	estimate	Lower CL	Upper CL	p.value
Intercept	23.506	21.787	25.224	0
slope	0.087	-0.071	0.245	0.273


Udbytte vs. Reaktionstal
udbytte
40
30
20
10
5.5
6.0
6.5
7.0
rt

Part 2: Analysis using propensity scores.

Our analysis is based on data from **an observational study**. In contrast to an experimental study, the "treatments" (e.g. the choice of variety, fertilizer, sowing depth) were not randomized. "Treatments" are all variables that the farmer can actively choose. In randomized studies we assume that confounding factors are randomly distributed across the treatment levels, this is not necessarily the case in an observational study. E.g. the choice of sowing depth might depend on the soil type. A farmer with field of a certain soil type might have a tendency (a high propensity) to choose a low sowing depth. Therefore, if we look at the effect of sowing depth on the yield, what we actually observe might be the effect of the soil type. The goal is to isolate the effect of the treatment (e.g. the choice of sowing depth) from the effects of the other factors. This can e.g. be done by reweighting the observations to obtain a more balanced data set. The reweighting is done by **propensity scores,** which give the probability of being assigned a certain treatment. We calculate propensity scores (via generalized boosted modeling) for the following variables ("treatments")

- Forforfrugt
- Forfrugt
- Husdyrgødning
- Andengødning
- Renbestand
- Sort
- Podemiddel
- Planteart_blandet
- Sådato
- Sådybde

and present the analysis for these variables in the tables below. For the majority of the variables we do not see a significant effect on the yield. The variables with a significant effect are the following:

- **Husdyrgødning:** Husdyrgødning has a significant positive effect on the yield (Table 3.c). We plot a histogram of "udbytte" colored by husdyrgødning for visualization.

- **Blandet_art:** The kind of crop that lupin is blend with has a significant effect on the yield. Vårhvede/Hvede has a higher yield than pure lupin or a mix with Vårtriticale (Table 3.j). This is the same result as in the analysis without propensity score weighting.

- **Sådybde**: Sowing Lupin at a medium sådybde (3.5cm – 4.0cm) gives a significantly higher yield than sowing "low" or "high". (Table 3.l)

- **Sådato:** Early sowing gives a significantly higher yield than late sowing. (Table 3.k)

In contrast to the analysis without propensity score weighting, we can no longer conclude that there is a positive significant effect of using blandsæd (compare Tabel 3e: renbestand). We can, however, conclude that there is a positive significant effect of blending lupin with Hvede/Vårhvede (compare Tabel 3.j: blandet_art). Blending with Vårtriticale seems to have a negative effect on the yield, there is, however, very few data about these blends. We plot a histogram of "udbytte" colored by the blend_art for visualization.

Tabel 3 a: Effect of forforfrugt – with propensity scores

Forforfrugt	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Bælg	23.648	21.457	25.838	a
Andet	23.774	21.388	26.16	a

Tabel 3 b: Effect of forfrugt – with propensity scores

Forfrugt	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Vår	22.337	19.496	25.178	a
Vinter	22.912	20.065	25.759	a
Andet	25.52	22.924	28.116	a

Tabel 3 c: Effect of husdyrgødning – with propensity scores

husdyrgødning	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Nej	23.107	20.632	25.582	a
Ja	27.016	24.537	29.495	b

Tabel 3 d: Effect of andengødning – with propensity scores

andengødning	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Nej	23.217	21.043	25.391	a
Ja	25.956	23.782	28.13	a

Tabel 3 e: Effect of renbestand – with propensity scores

Renbestand	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Renbestand	22.887	20.631	25.144	a
Blandsæd	26.012	23.762	28.263	a

Tabel 3 h: Effect of sort – with propensity scores

sort	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Bazalt	15.75	7.261	24.239	a
Bolero	20.308	14.523	26.093	a
Zeus	21.977	16.411	27.542	a
Tango	22.471	17.618	27.323	a
Primadonna	23.152	18.804	27.499	a
Regent	23.624	18.095	29.152	a
Andet	24.152	18.475	29.828	a
Rumba	24.98	20.822	29.138	a
Iris	25.253	21.064	29.442	a
Borgine	29.135	24.982	33.288	a

Tabel 3 i: Effect of podemiddel – with propensity scores

Podemiddel	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Ingen	21.285	13.874	28.696	a
Legumino	21.53	8.755	34.305	a
LegumeFix	23.455	21.289	25.621	a
Andet	24.491	21.676	27.305	a

Tabel 3 j: Effect of blandet_art– with propensity scores

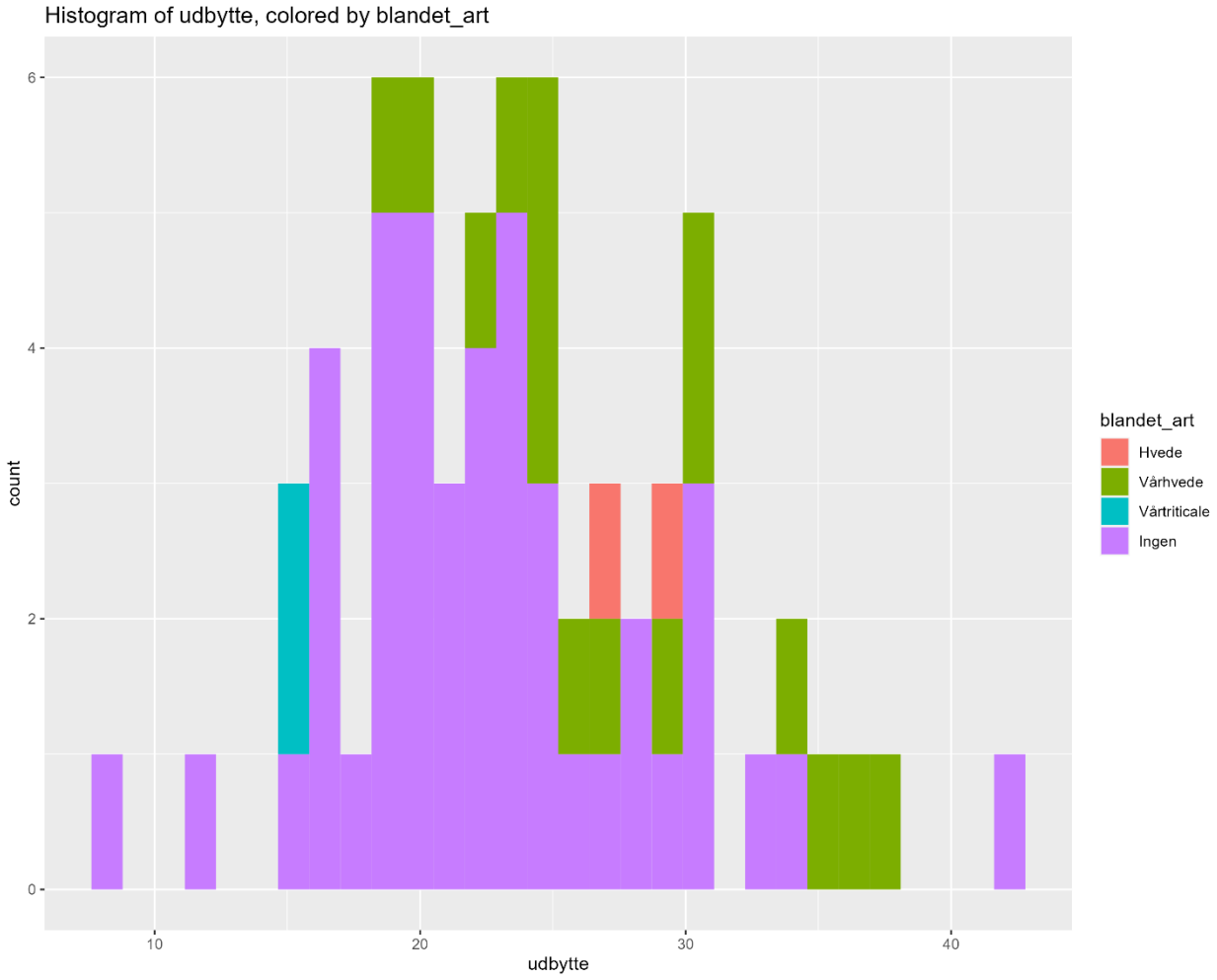
Blandet_art	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Vårtriticale	15	7.734	22.266	a
Ingen	22.66	20.85	24.47	a
Vårhvede	27.851	25.176	30.526	b
Hvede	28	20.611	35.389	ab

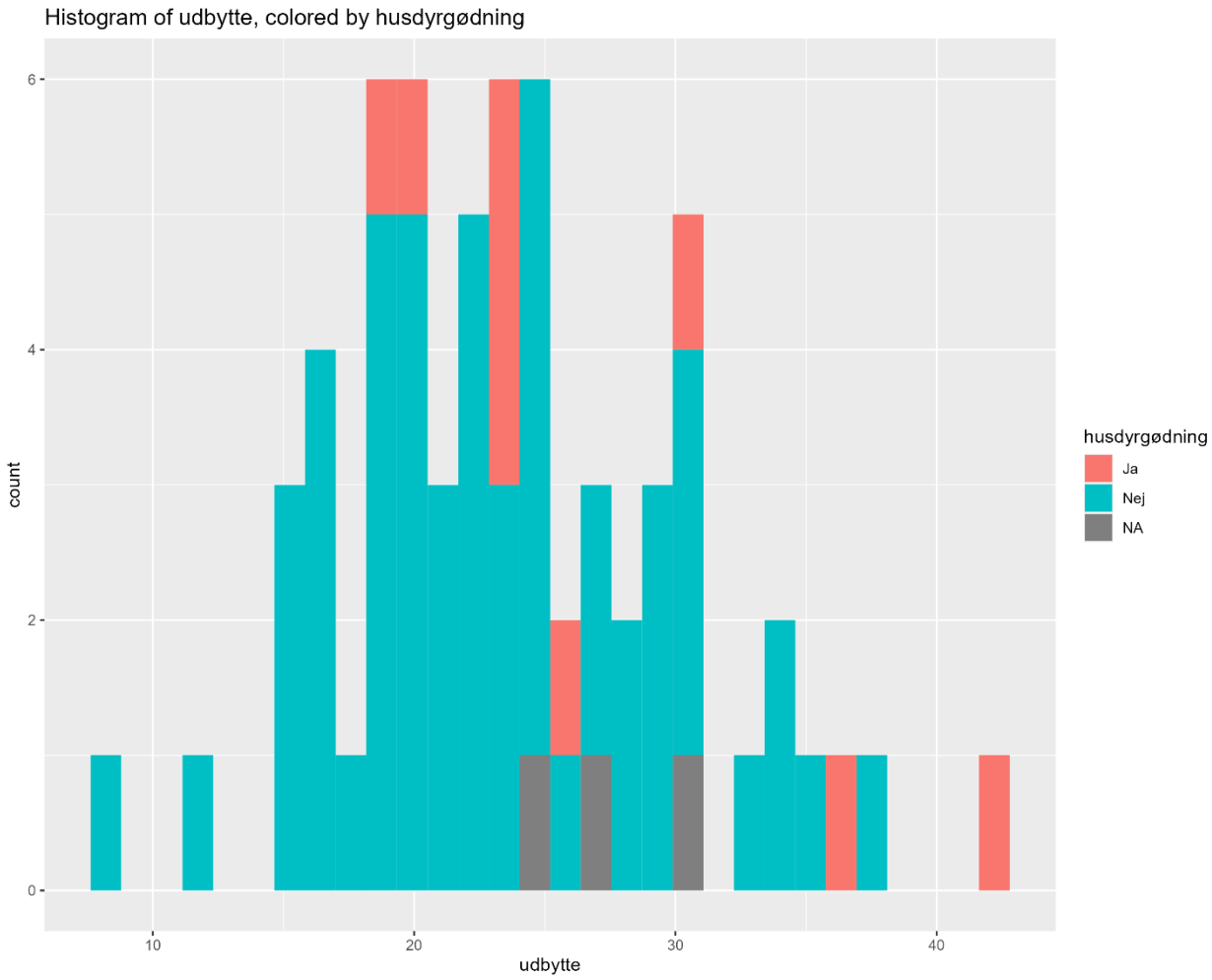
Tabel 3 k: Effect of sådato– with propensity scores

Sådato	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Late	21.388	19.403	23.373	a
Early	26.875	24.88	28.869	b

Tabel 3 l: Effect of sådybde– with propensity scores

Sådybde	udbytte	Lower CL	Upper CL	Signifikans Gruppe
Low	21.382	18.94	23.825	a
High	23.92	21.577	26.263	a
Medium	28.668	25.973	31.363	b


Histogram of udbytte, colored by blandet_art
count
6
4
2
0
10
20
30
40
udbytte
blandet_art
Hvede
Vårhvede
Vårtriticale
Ingen


Histogram of udbytte, colored by husdyrgødning
count
udbytte
0
2
4
6
10
20
30
40
husdyrgødning
Ja
Nej
NA