Hailstorm Analysis Report

Dennis Chomba

2025-07-06

Table of Contents

0.1 🔲 Data Import & Cleaning

Data was cleaned to deal with missing values and outliers for easier analysis.

```
hail_data <- read_excel("C:\\Users\\User\\Desktop\\Copy of HAILSTORM_IMPACT_D
ATA(1).xlsx")

hail_data <- hail_data %>%
  rename(
    Girth1 = `...4`,
    Girth2 = `...5`,
    Girth3 = `...6`,
    Flower_Bud = `FLOWER/BUD`
  ) %>%
  mutate(
    across(c(PLOT_NO., VEGETATIVE, Girth1, Girth2, Girth3, `Girth size (mm)`,
Flower_Bud), as.numeric)
```

```
) %>%
  filter(!is.na(PLOT_NO.))
```

0.2 ⊞ Exploratory Visualizations

Interpretation

1. The box plot below showes that the majority of girth measurements clustered tightly around ~2 mm, with very low variation. A few outliers extended above 8 mm, but they were rare.

2. This lack of variability limits girth size's usefulness because most plants had similar stem thickness, it's difficult to assess how much girth impacts bud abortion in this dataset.

0.2.1 Girth Size Distribution

```
ggplot(hail_data, aes(y = `Girth size (mm)`)) +
  geom_boxplot(fill = "orange", alpha = 0.6) +
  labs(title = "Distribution of Girth Size (mm)")
```


Distribution of Girth Size (mm)
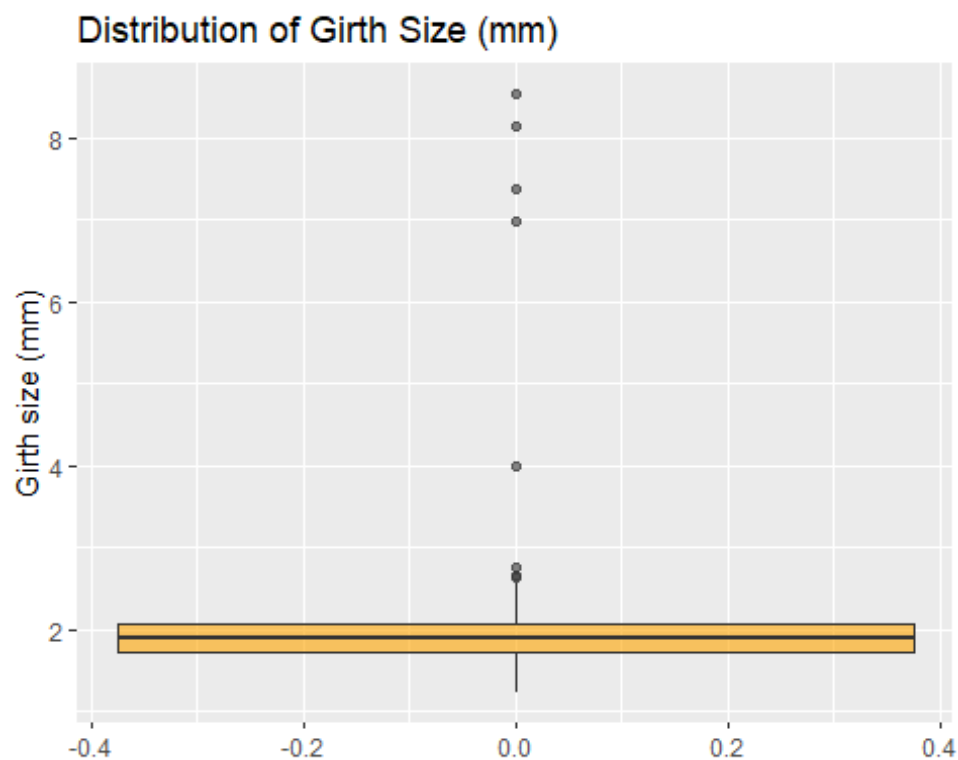
0.2.2 Flower/Bud Abortion Distribution

Interpretation

The abortion percentage displayed a wide, symmetrical distribution ranging from below 20% to as high as 97%, with a median around 50%.

From an analytical perspective—it gives you enough spread in our outcome variable (Flower_Bud) to reveal meaningful patterns and relationships. It also highlights the varying levels of hailstorm impact across plots

```
ggplot(hail_data, aes(y = Flower_Bud)) +
  geom_boxplot(fill = "tomato", alpha = 0.6) +
  labs(title = "Distribution of Flower/Bud Abortion (%)")
```


Distribution of Flower/Bud Abortion (%)
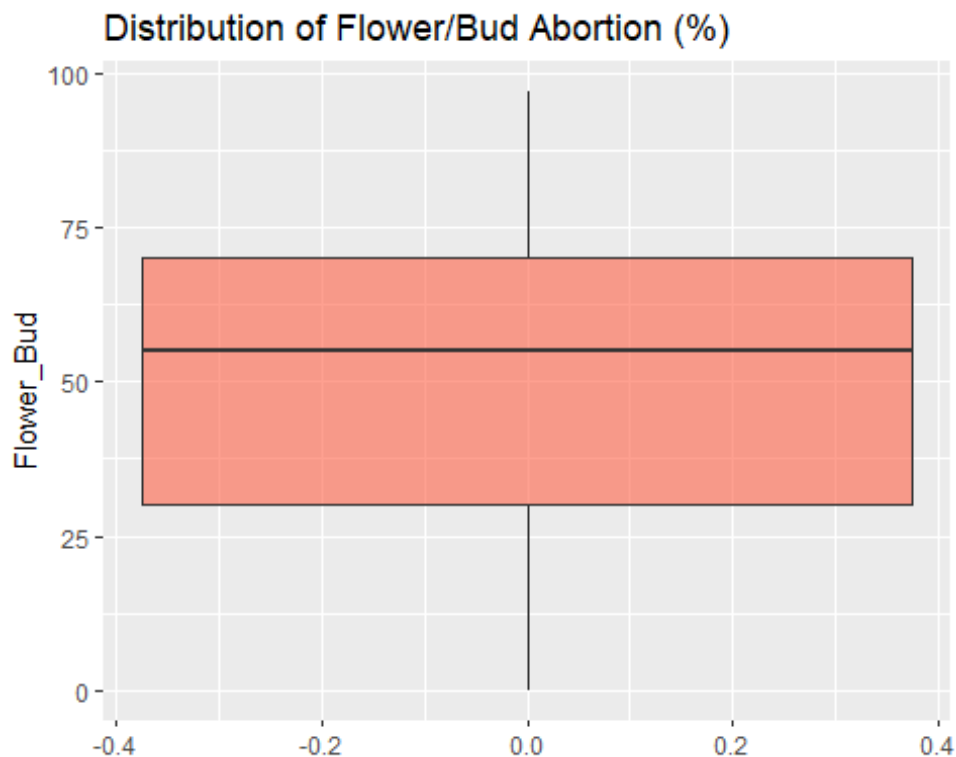
0.2.3 Stem Strength Distribution

Interpretation

Stem scores ranged from 1 to 3 with a slightly right-skewed distribution. Most scores clustered around 1.5–2.0, with few plants scoring the maximum of 3.

This means there's moderate variability suggesting that higher stem scores are associated with higher abortion rates, hinting at a potential resource loss/ low /no yield.

```
ggplot(hail_data, aes(y = STEM)) +
  geom_boxplot(fill = "skyblue", alpha = 0.6) +
  labs(title = "Distribution of Stem Strength")
```

Distribution of Stem Strength
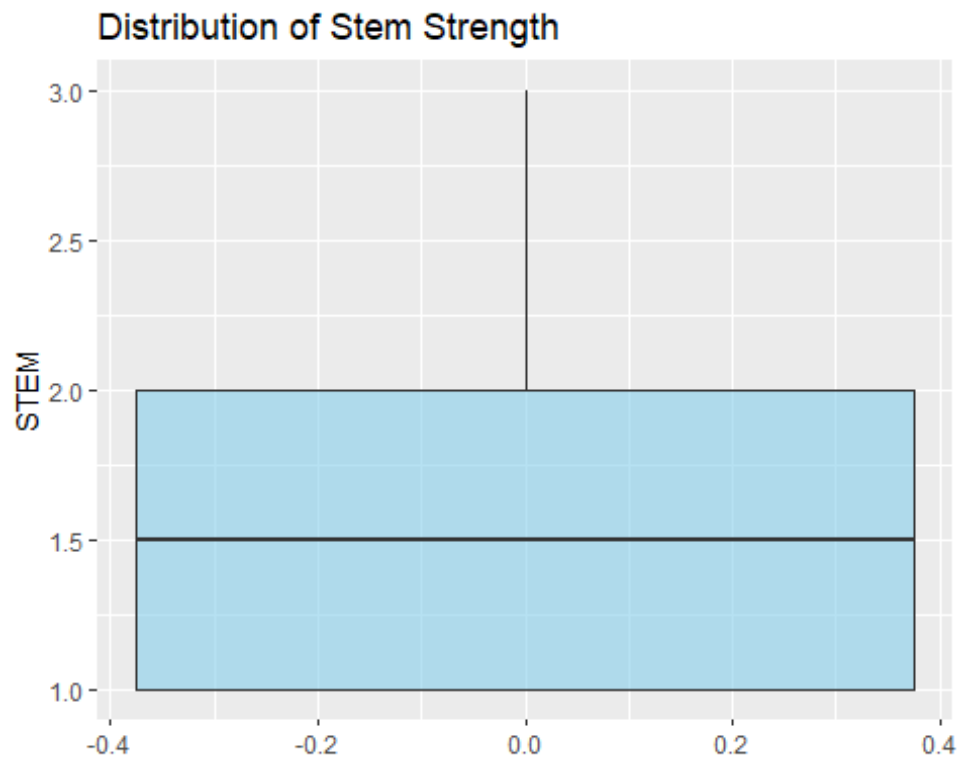
0.3 ⬚ Correlation Analysis

Interpretation.

VEGETATIVE ↔ Flower_Bud: 0.38 (moderate positive)

STEM ↔ Flower_Bud: 0.33 (moderate positive)

Girth size ↔ Flower_Bud: –0.06 (no correlation)

This shows that more vegetative and structural growth is positively correlated with bud abortion. This could reflect a stress-induced imbalance in how resources are distributed during recovery from hailstorms. Girth size's negligible correlation supports its low importance in further analyses.

```
numeric_data <- hail_data %>%
  select(STEM, VEGETATIVE, Flower_Bud, `Girth size (mm)`) %>%
  na.omit()

cor_matrix <- cor(numeric_data)
round(cor_matrix, 2)
```

```
##                STEM VEGETATIVE Flower_Bud Girth size (mm)
## STEM           1.00      0.43        0.33              0.07
## VEGETATIVE     0.43      1.00        0.38              0.00
## Flower_Bud     0.33      0.38        1.00             -0.06
## Girth size (mm) 0.07     0.00       -0.06              1.00
```

0.4 ▣ Relationship Visualizations

Interpretation

The plot shows a very weak negative relationship. While there's a trend where plants with thicker stems tend to have slightly lower abortion rates.

This visual matches the correlation and model results. Girth alone doesn't explain much variation in abortion rate—likely due to its limited range.

0.4.1 Flower Abortion vs Girth Size

```
ggplot(hail_data, aes(x = `Girth size (mm)`, y = Flower_Bud)) +
  geom_point(color = "darkred", alpha = 0.7) +
  geom_smooth(method = "lm", se = FALSE, color = "black") +
  labs(title = "Flower Abortion vs Girth Size", x = "Girth (mm)", y = "Flower
/Bud Abortion (%)")
```
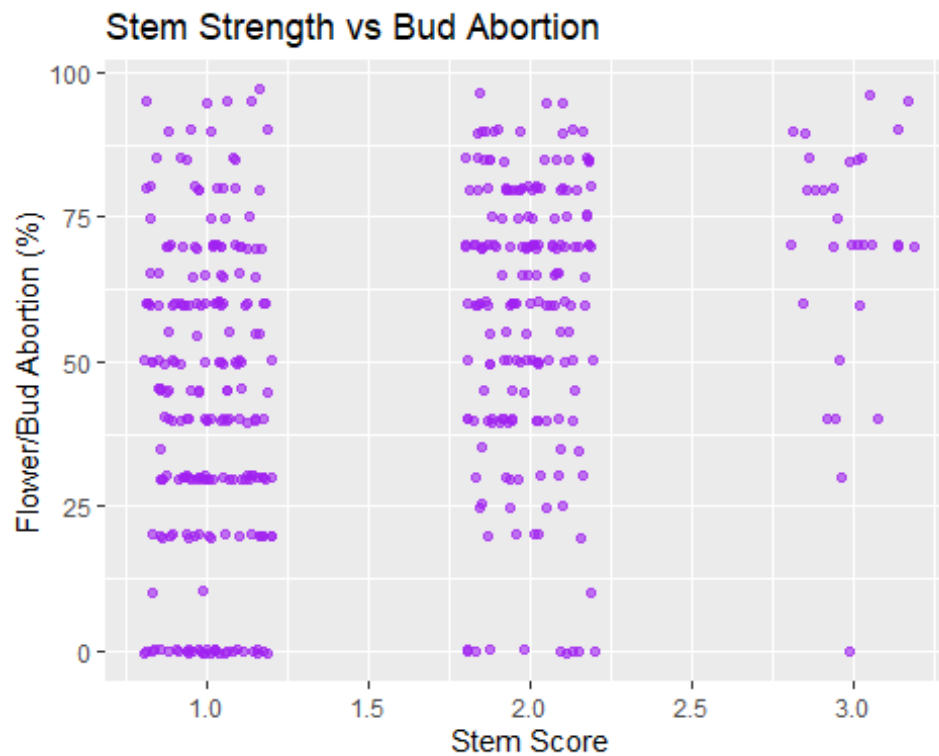

Flower Abortion vs Girth Size
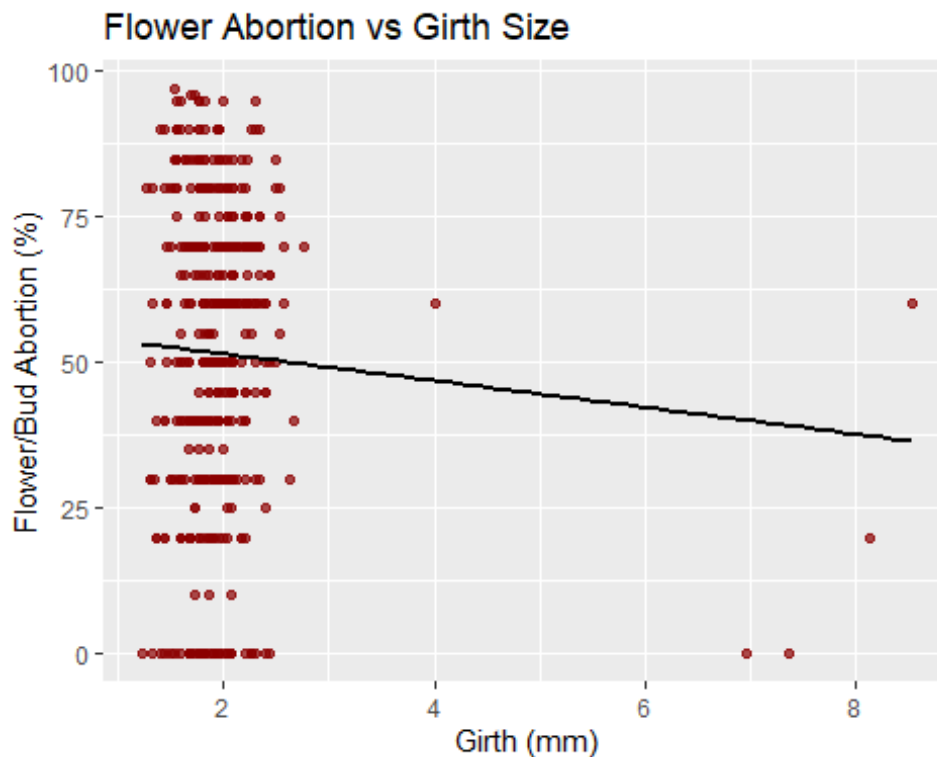
0.4.2 Stem Strength vs Flower Abortion

Interpretation

This shows that there is a strong impact on bud abortion.

```
ggplot(hail_data, aes(x = STEM, y = Flower_Bud)) +
  geom_jitter(width = 0.2, color = "purple", alpha = 0.6) +
  labs(title = "Stem Strength vs Bud Abortion", x = "Stem Score", y = "Flower
/Bud Abortion (%)")
```


0.5   ⬛ Linear Model: Predicting Flower Abortion

Interpretation

Model R-squared: 0.186 → 18.6% of variation explained

Significant predictors:

VEGETATIVE (p < 0.001, β = +4.53)

STEM (p < 0.001, β = +8.97)

Non-significant predictor:

Girth size (mm) (p = 0.117, β = –2.87)

Our model confirms that stem strength and vegetative growth are significantly associated with higher bud abortion. While this may seem counterintuitive, it suggests that under extreme hail stress, plants that prioritize regrowth (structure or foliage) might sacrifice reproductive organs like buds and flowers. Girth plays a very minor, non-significant role, aligning with earlier insights.

```
model <- lm(Flower_Bud ~ `Girth size (mm)` + STEM + VEGETATIVE, data = hail_d
ata)
summary(model)

##
## Call:
## lm(formula = Flower_Bud ~ `Girth size (mm)` + STEM + VEGETATIVE,
##      data = hail_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -68.032 -14.028   4.703  16.875  55.899
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        18.1300     5.4283   3.340 0.000912 ***
## `Girth size (mm)`  -2.8686     1.8265  -1.571 0.117023
## STEM                8.9696     2.0503   4.375 1.53e-05 ***
## VEGETATIVE          4.5332     0.7492   6.051 3.15e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 23.93 on 428 degrees of freedom
## Multiple R-squared:  0.186,  Adjusted R-squared:  0.1803
## F-statistic: 32.61 on 3 and 428 DF,  p-value: < 2.2e-16
```

0.6   Top 10 Most Affected Plots

Interpretation

The most severely affected plots all had abortion rates of 95% or higher. Most of these plots:

Had high vegetative scores (7.0–9.5)

Had low to moderate stem strength (mostly 1–2)

Had narrow girth (1.5–2.3 mm)

These plots could represent vulnerable phenotypes. They are priority targets for intervention or follow-up study. Despite having high vegetative growth, their reproductive systems failed—possibly suggesting these plants diverted energy toward regrowth rather than protecting developing buds.

```
top_affected <- hail_data %>%
  arrange(desc(Flower_Bud)) %>%
  slice_head(n = 10)

knitr::kable(top_affected)
```

PLOT_NO.	VEGETATIVE	STEM	Girth1	Girth2	Girth3	Girth size (mm)	Flower_Bud
348	7.0	1	1.5	1.8	1.3	1.533333	97
345	9.0	3	2.0	1.3	1.8	1.700000	96
351	8.0	2	1.4	1.8	2.0	1.733333	96
191	9.5	1	2.0	2.3	2.6	2.300000	95
324	7.0	1	1.8	2.2	1.5	1.833333	95
334	9.0	2	1.6	1.4	1.7	1.566667	95
339	7.0	1	1.4	1.5	1.9	1.600000	95
340	7.0	2	1.5	2.0	1.8	1.766667	95
364	7.0	1	1.8	1.5	2.0	1.766667	95
378	9.0	3	2.1	1.6	2.3	2.000000	95

0.7 ▢ Combined Dashboard

Interpretation

The results below is a summary of the distribution across the plot.

```
p1 <- ggplot(hail_data, aes(x = PLOT_NO., y = Flower_Bud)) +
  geom_col(fill = "tomato") +
  labs(title = "Flower Abortion per Plot", x = "Plot", y = "% Abortion") +
  theme_minimal()

p2 <- ggplot(hail_data, aes(x = PLOT_NO., y = STEM)) +
  geom_col(fill = "forestgreen") +
  labs(title = "Stem Strength per Plot", x = "Plot", y = "Stem Score") +
  theme_minimal()

p3 <- ggplot(hail_data, aes(x = PLOT_NO., y = `Girth size (mm)`)) +
  geom_col(fill = "steelblue") +
  labs(title = "Girth Size per Plot", x = "Plot", y = "Girth (mm)") +
  theme_minimal()

(p1 / p2 / p3)
```

Flower Abortion per Plot

Stem Strength per Plot

Girth Size per Plot
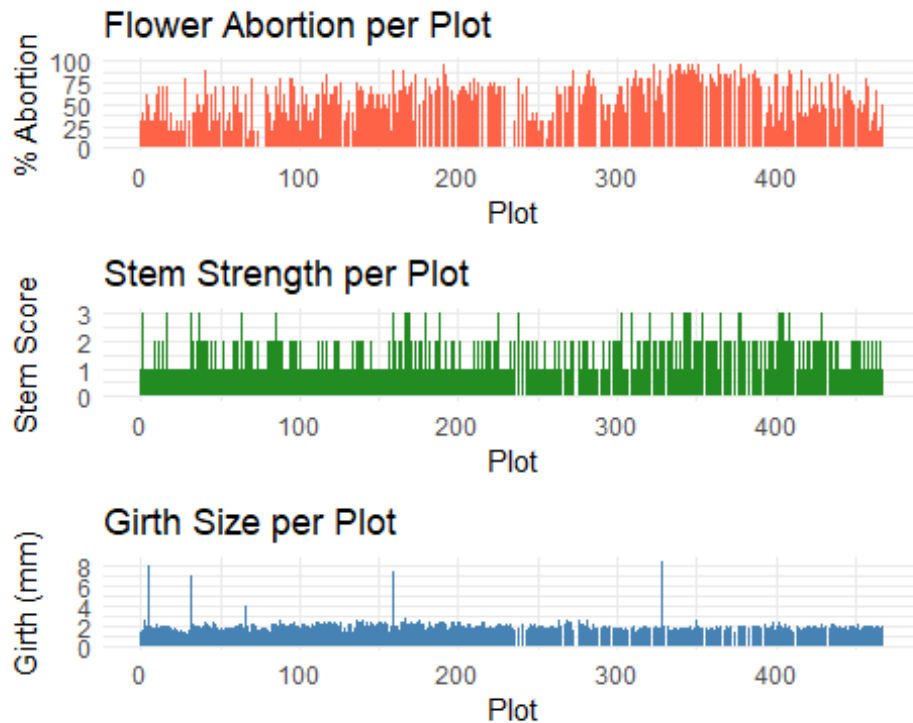
challenges

1. Limited Sample Size

2. Missing Agronomic or Environmental Context Variables like soil quality, rainfall, hailstorm intensity per plot, or flowering stage were not included.

3. Imbalanced Trait Distribution Certain traits (e.g. Girth size) had very low variability, while others (e.g. Flower_Bud) were highly dispersed.

Conclusion.

This analysis provides critical insights into the physiological responses of plants subjected to hailstorm damage. The investigation revealed that flower/bud abortion rates varied significantly across plots, with the most affected plots experiencing abortion rates exceeding 95%. Notably, vegetative growth and stem strength emerged as the most significant predictors of abortion, both displaying positive associations with reproductive loss. This suggests a possible trade-off in stress recovery, where energy invested in vegetative and structural regrowth may reduce the plant's capacity to maintain reproductive tissues.

In contrast, girth size (mm) showed minimal variation and was not significantly associated with flower abortion, indicating its limited role in predicting damage severity under these conditions.

The identification of highly affected plots and vulnerable growth traits can inform future agronomic strategies. By targeting plots with high vegetative scores and weak structural integrity, intervention efforts can be better prioritized to mitigate crop loss. These findings underscore the importance of integrated plant monitoring post-hailstorm to balance structural resilience and reproductive performance.

Recommendations.

1. Prioritize Plots with High Bud Abortion for Intervention. Plots with >95% flower/bud abortion should be flagged for immediate suppor or recovery measures. Replanting, supplementary feeding, or stress recovery protocols.

2. Breeding Selection Toward Balanced Vigor. Since both VEGETATIVE and STEM scores were positively correlated with flower abortion. Develop or select genotypes with moderate vegetative traits and well-distributed structural strength, optimizing for both growth and floral retention.

3. Deploy Stress Monitoring Post-Hailstorm Use real-time monitoring of vegetative recovery and flower drop and field scoring sheets for STEM and VEGETATIVE to proactively estimate reproductive loss risk. This approach enables faster diagnosis and prioritization even before full data analysis is completed.