# Agricultural Dataset of an Argentinian Corn Field

Yield monitor data for a corn field in Argentina with variable nitrogen.

Jack Nguyen, Nikita Kiselov

2025-12-18

## 1 Introduction

### 1.1 Dataset Overview

This dataset covers the experimental study of corn yield and nitrogen fertilizer treatment with field characteristics from the Las Rosas farm in Rio Cuarto, Cordoba, Argentina done by Anselin, Bongiovanni, and Lowenberg-DeBoer (2004). The data was collected using yield monitor from strip trials during the harvests in years 1999 and 2001 to determine the potential of spatial econometric analysis in estimating the site-specific crop response functions when it comes to applying nitrogen treatment. There are 3443 operations and 8 variables, including the year, latitude, longitude, yield in quintals/ha, amount of nitrogen in kg/ha, topographic factor, brightness value, rep factor, and nitrogen as a factor. Of these variables, both the topographic, rep, and nitrogen factors, are qualitative while the rest are quantitative.

### 1.2 Summary Statistics

For each nitrogen level, the mean was obtained by dividing the sum of all yield values by the number of observations in the sample, while the median was obtained by finding the middle value from (the sorted version of) that sample.
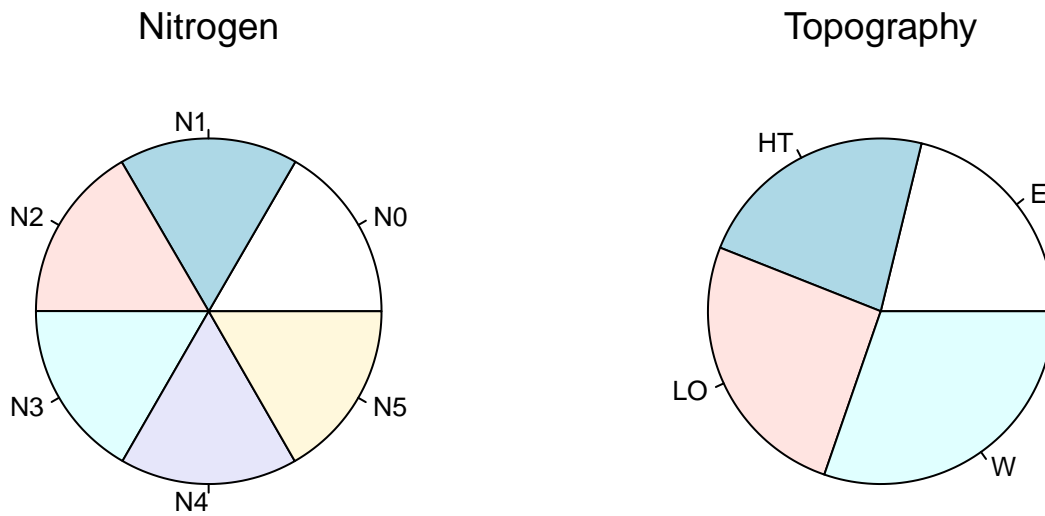
The standard deviation was calculated as the square root of the sum of squared differences, between each yield and mean, divided by the result of subtracting sample size by one. The interquartile range (IQR) was obtained by finding the lower and upper medians of the two halves of the sample data that is by the median, and then subtracting the lower median from the upper one. The range was identified by simply obtaining the lowest and highest values from the sample. The mode was not needed in this study and therefore was never calculated for each level.

Also for each nitrogen level, the crops were given approximately the same amount of nitrogen treatment. However, this is not the case in terms of topography levels, as each area of these levels are different. West Slope (W) recieved the most amount of nitrogen treatment, followed by Low East (LO) and Hilltop (HT), while East Slope (E) recieved the least.

Table 1: Summary Statistics of Yield (quintals/ha) by Nitrogen Fertilizer Level

| Level | Size | Mean | Median | SD | IQR | Range |
|-------|------|------|--------|------|------|----------------|
| N0 | 573 | 65.0 | 62.0 | 20.9 | 36.0 | 12.66 - 108.84 |
| N1 | 577 | 68.6 | 65.2 | 19.2 | 29.8 | 27.44 - 110.54 |
| N2 | 571 | 69.7 | 67.1 | 19.3 | 28.1 | 31.79 - 112.85 |
| N3 | 575 | 70.3 | 66.8 | 19.2 | 27.7 | 19.41 - 110.12 |
| N4 | 572 | 72.6 | 69.2 | 19.1 | 27.5 | 32.05 - 117.9 |
| N5 | 575 | 72.8 | 70.3 | 20.1 | 29.8 | 31.79 - 117.19 |

Figure 1: Distribution of nitrogen and topography treatment levels

## Nitrogen

## Topography



## 1.3 Farm topography and yield

In the order from the largest area to the smallest, these topography levels are Low East (LO), Hilltop (HT), East Slope (E), and West Slope (W).
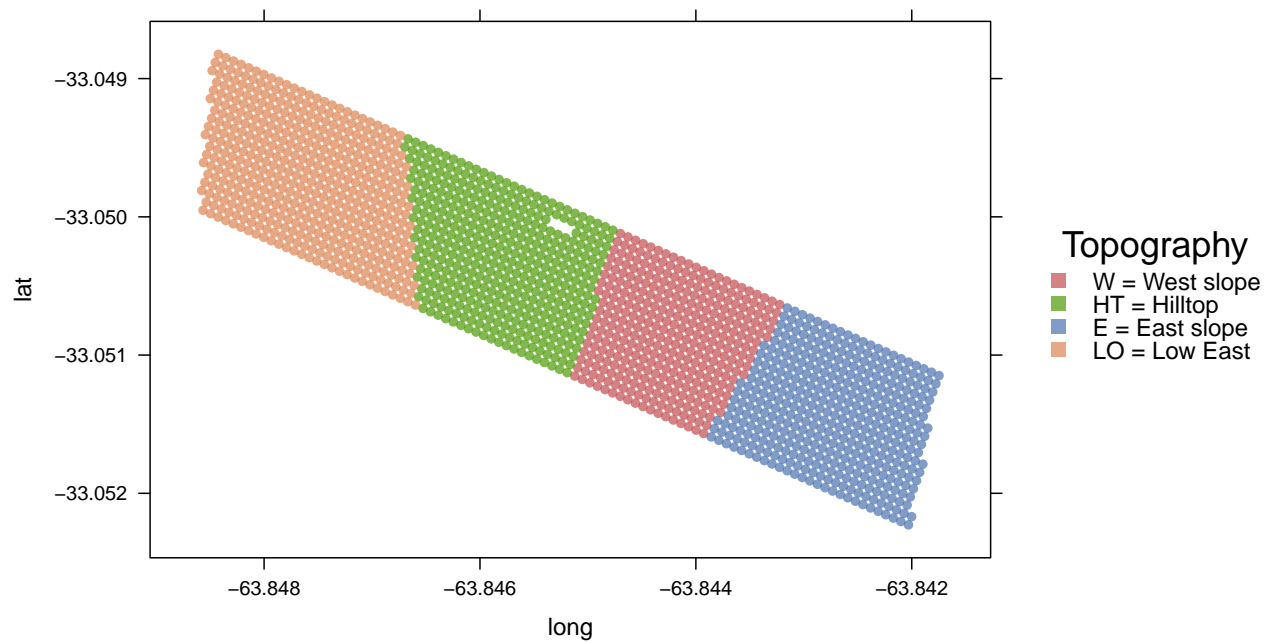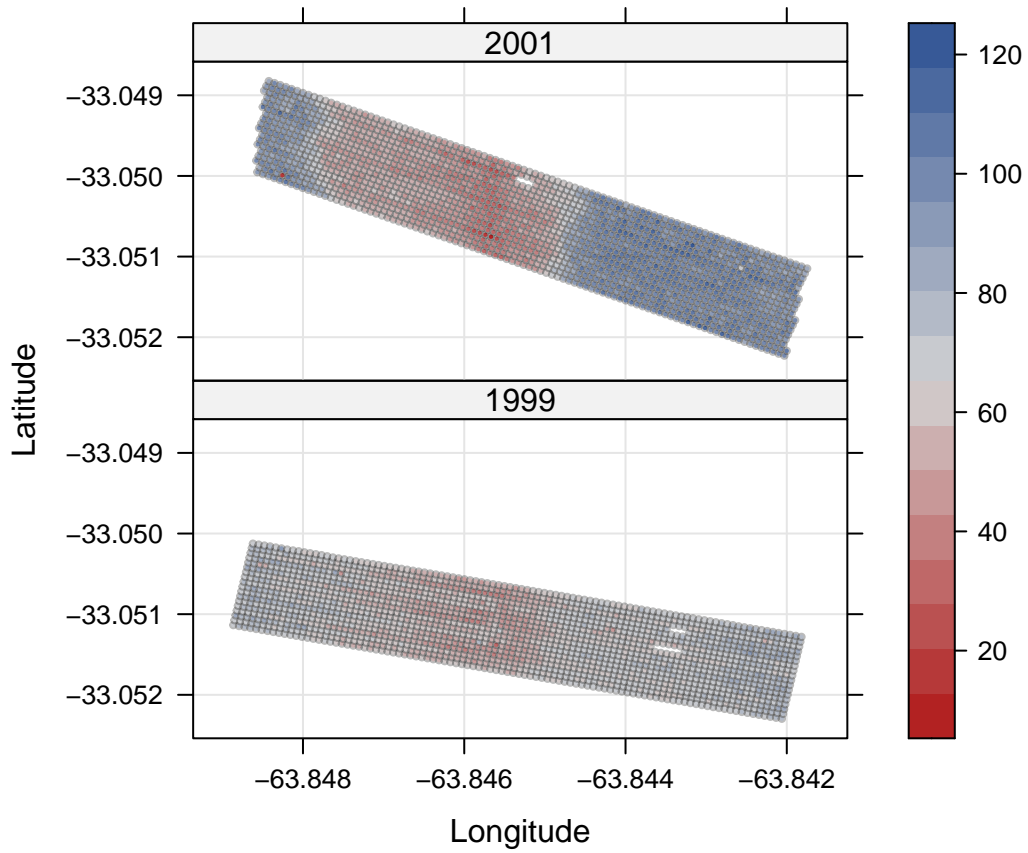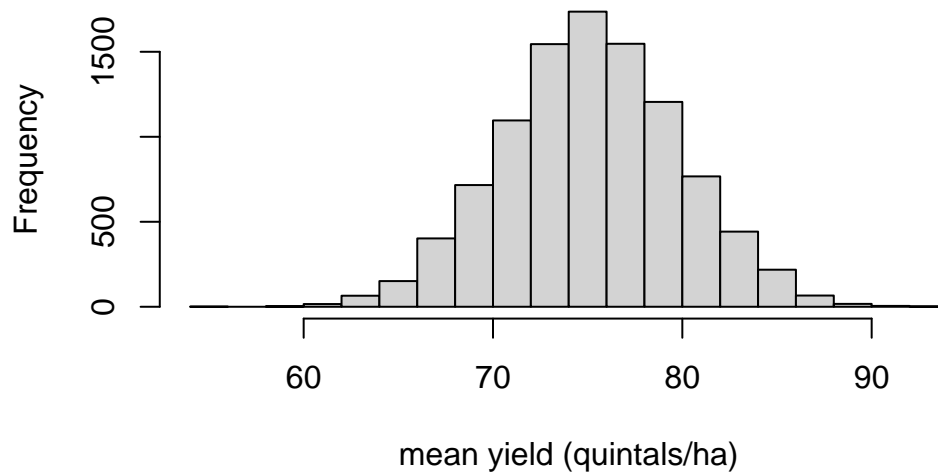
Figure 2: Farm topography in 2001

Figure 3: Corn yield distribution by location



## 1.4 Sampling distribution

Figure 4: Sampling distribution of mean corn yield in 2001 for 10000 samples of size 30



As seen in Figure 4, the sampling distribution is symmetric around 75 and approximately normal.
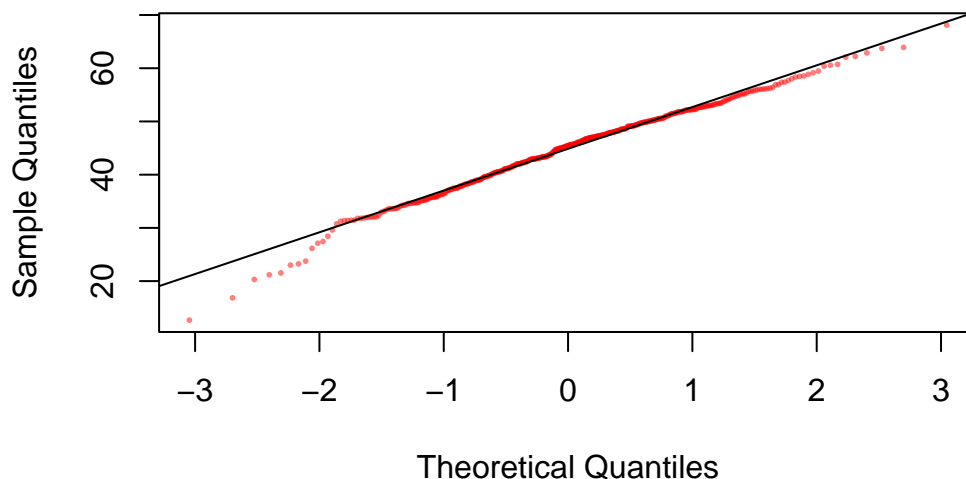
# 2 One Parameter Test

The mean corn yield in 2001 across all topographies was 75.2 quintals/ha. However, as seen in Figure 3, the corn yield differed between various parts of the farm. To test whether topography affects corn growth, we compare the true mean corn yield on hilltop topography ($\mu$) to the mean yield of the entire farm in 2001. We use one-sample t-test with the following hypotheses:

$$H_0 : \mu = 75.2 \quad \text{(same as farm mean)}$$

$$H_a : \mu \neq 75.2 \quad \text{(different from farm mean)}$$

There are two assumptions to verify: normality and independence. Figure 5 shows that the yield distribution is very close to normal apart from a few outliers on the lower end. However, this is expected since some plants could have failing yield due to parasites or other environmental factors.

Figure 5: Q-Q plot for corn yield on hilltop topography in 2001



Independence can't be fully claimed since the data was collected from a single field. Therefore the result can't be generalized to corn yield on hilltop topography in general. However, if the scope of this analysis is limited to the Las Rosas farm and nearby area, the sample contains enough independent measurements and is representative, so the test can continue.

Table 2: t-test results for hilltop true mean yield

| p-value | df | 99% CI |
|---------|-----|--------------|
| 6.2e-258 | 430 | 43.68 - 45.69 |

Based on the data from Las Rosas farm in 2001, we are 99% confident that the true mean yield of corn on hilltop topography of the farmland is between 43.68 and 45.69. Since this interval does not contain 75.2, there is statistically significant evidence against the null hypothesis at 1% significance level. We conclude that the true mean corn yield on hilltop topography is different from the mean yield of all topographies in the Las Rosas farm.

## 2.1 Advanced Tables with kableExtra

For even more formatting options, use the {kableExtra} package:

Table 3: Advanced table formatting with kableExtra

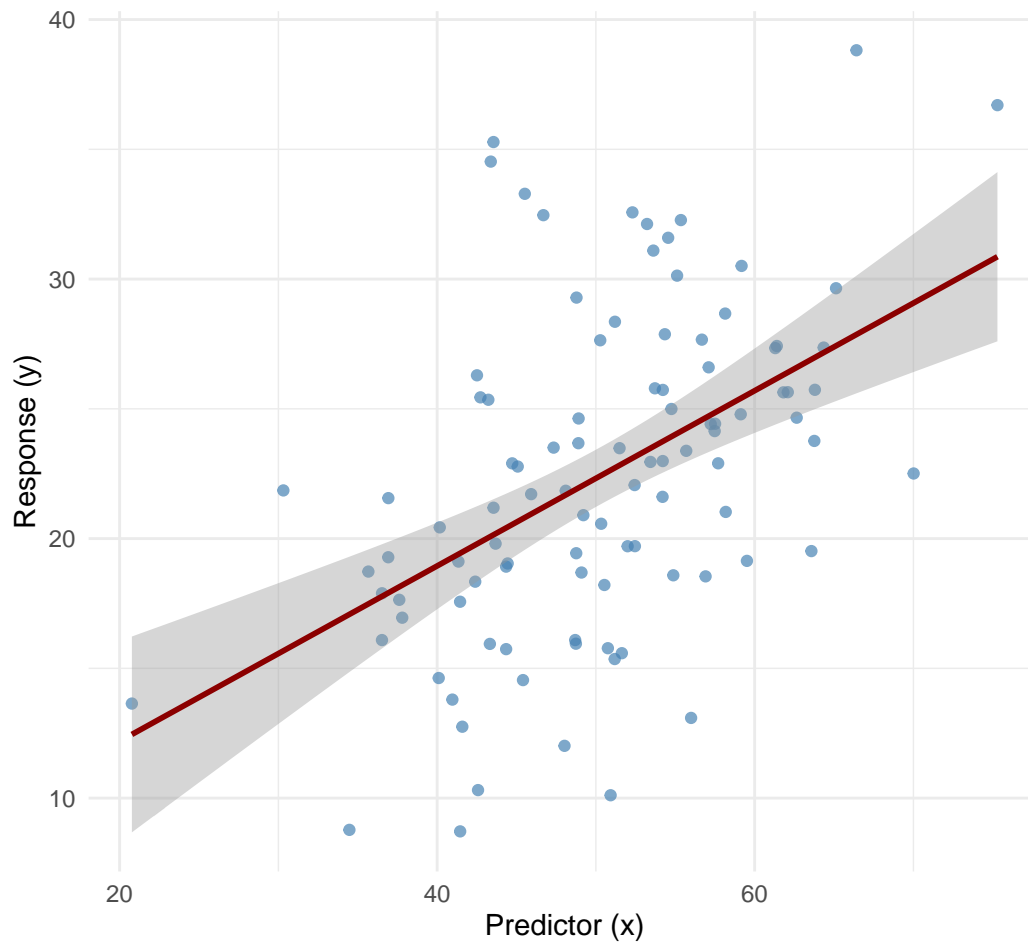| X Category | Count | Mean Y | SD Y | Min Y | Max Y |
|---|---|---|---|---|---|
| Low | 10 | 17.24 | 3.84 | 8.78 | 21.86 |
| Medium | 68 | 21.96 | 6.40 | 8.72 | 35.28 |
| High | 22 | 26.13 | 4.77 | 19.14 | 38.82 |

Notice how Table 3 includes footnotes and styling. The `html-table-processing: none` option prevents Quarto from re-processing tables created with `{kableExtra}`.

# 3 Creating Figures with R

Figures generated from R code should include proper labels, captions, and alt text for accessibility.

## 3.1 Basic Plot with Caption

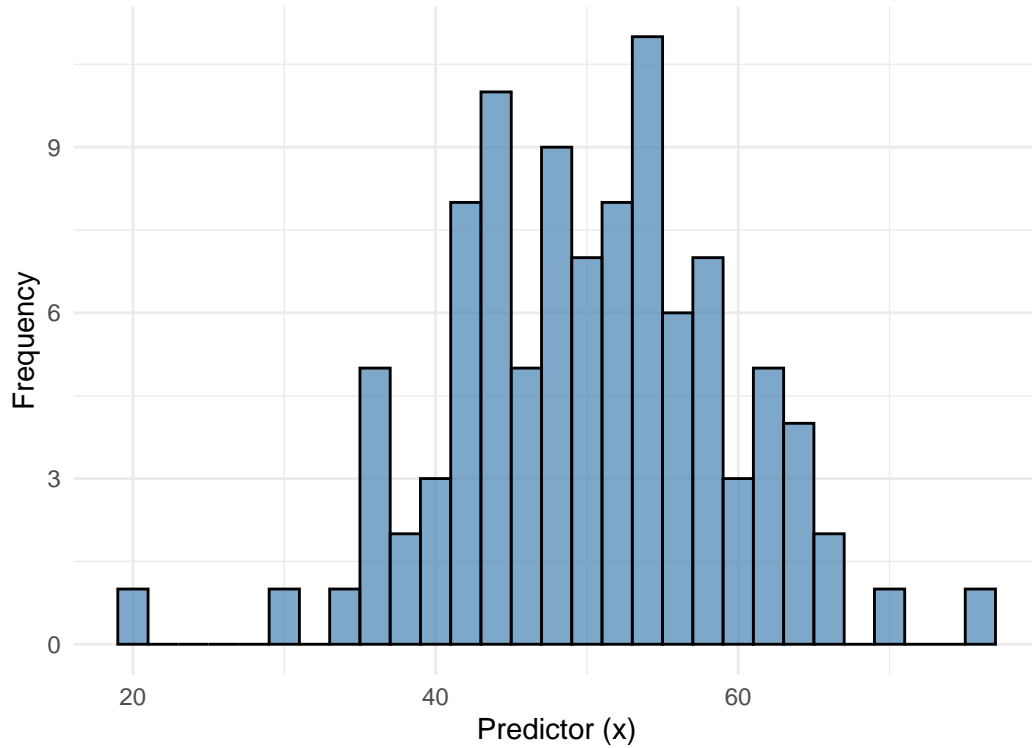Figure 6: Relationship between predictor x and response y



As shown in Figure 6, there appears to be a positive linear association between the predictor and the response.
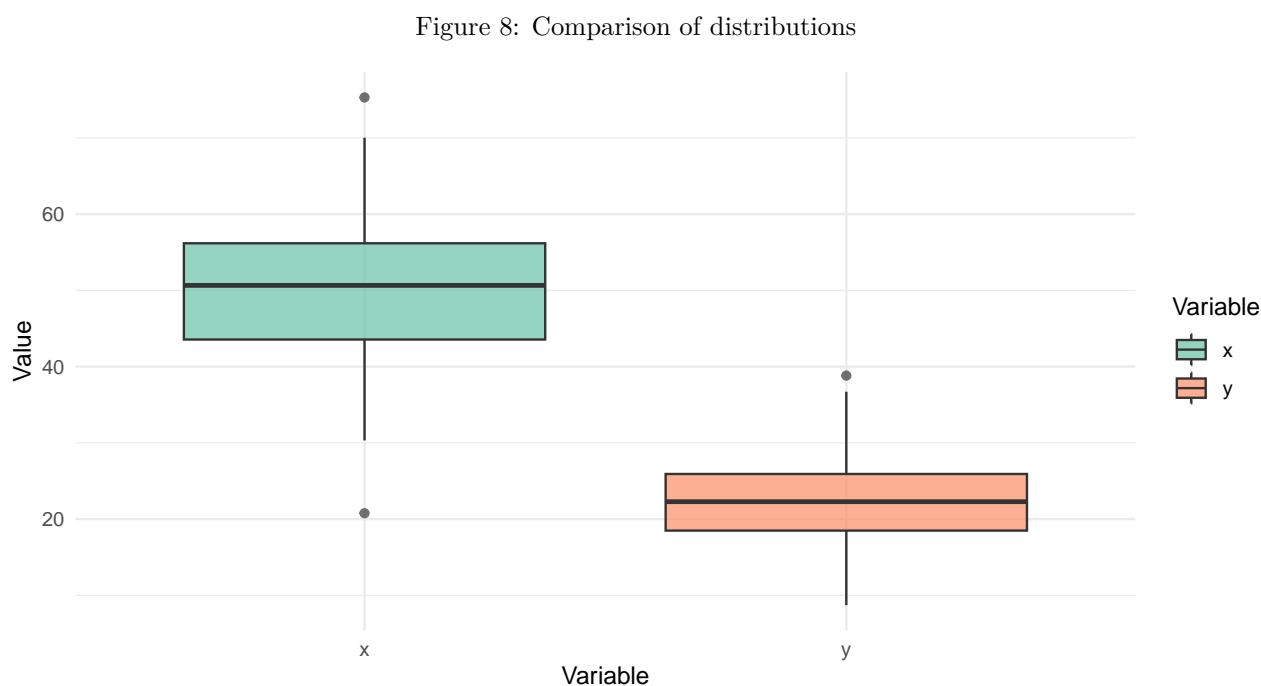
## 3.2 Plot with Alt Text from Code

You can also specify alt text within the `labs()` function and extract it automatically:

Figure 7: Distribution of predictor variable x



The distribution in Figure 7 shows that x values are approximately normally distributed.

## 3.3 Multiple Plots

Figure 8: Comparison of distributions



# 4 Including External Images

To include external images (not generated by R), you can use markdown syntax:

```
![Caption for your image](path/to/image.png){#fig-external}
```

Or use knitr:

```
knitr::include_graphics("path/to/image.png")
```

# 5 Cross-Referencing

Cross-referencing allows you to refer to tables, figures, and sections by name rather than hard-coding numbers.

## 5.1 Referencing Tables

To reference a table, use `@tbl-label`. For example:

- See **?@tbl-kable-example** for summary statistics
- Compare results in Table 3

## 5.2 Referencing Figures

To reference a figure, use `@fig-label`. For example:

- The scatterplot in Figure 6 shows a clear trend
- Distribution details are in Figure 7

## 5.3 Referencing Sections

You can also reference sections by adding labels to headings. For example, the heading:

```
# Data Analysis {#sec-analysis}
```

Can be referenced as: See **?@sec-analysis** for details.

# 6 Key Points to Remember

1. **Table Labels**: Must start with `tbl-` (e.g., `#| label: tbl-results`)
2. **Figure Labels**: Must start with `fig-` (e.g., `#| label: fig-plot1`)
3. **Section Labels**: Use `{#sec-name}` after heading text
4. **Alt Text**: Always include `fig-alt` for accessibility
5. **Captions**: Use `fig-cap` for figures and `tbl-cap` for tables
6. **Code Appendix**: Automatically collects all your code at the end

# 7 Formatting Tips

- Use **bold** text with `**text**` or `__text__`
- Use *italic* text with `*text*` or `_text_`
- Create lists with `-` or numbered lists with `1.`, `2.`, etc.
- Add inline math with `$equation$` (e.g., $\mu = 50$, $\sigma^2 = 100$)
- Add display math with `$$equation$$`

# 8 Conclusion

This template provides a foundation for creating professional statistical reports using Quarto. Key features include:

- Professional tables using both markdown and `kable()`
- Figures with proper captions and alt text for accessibility
- Cross-referencing capabilities for tables, figures, and sections
- Automatic code appendix generation

When you start your own project, replace this content with your analysis while maintaining the structure demonstrated here.

# 9    References

Anselin, Luc, Rodolfo Bongiovanni, and Jess Lowenberg-DeBoer. 2004. "A Spatial Econometric Approach
    to the Economics of Site-Specific Nitrogen Management in Corn Production." *American Journal of
    Agricultural Economics* 86 (3): 675–87. https://doi.org/https://doi.org/10.1111/j.0002-9092.2004.00610.
    x.

# 10    Code Appendix

The code below shows all the R code used to generate the tables and figures in this document. The code is
displayed here for transparency and reproducibility but is hidden in the main body of the report.

```r
library(tidyverse)
library(knitr)
library(kableExtra)
library(agridat)
data(lasrosas.corn)
dat <- lasrosas.corn

sample_data <- tibble(
  id = 1:100,
  x = rnorm(100, mean = 50, sd = 10),
  y = 3 + 0.4 * x + rnorm(100, sd = 5)
)

dat %>% group_by(nf) %>%
  summarise(
    N = n(),
    Mean = mean(yield),
    Median = median(yield),
    SD = sd(yield),
    IQR = IQR(yield),
    Range = paste(format(min(yield), nsmall = 1), "-", format(max(yield), nsmall = 1)),
    .groups = "drop"
  ) %>% kable(
      digits = 1,
      col.names = c("Level", "Size", "Mean", "Median", "SD", "IQR", "Range"))
par(mfrow = c(1, 2))

pie(table(dat$nf), main = "Nitrogen", radius = 0.9, cex = 0.8, cex.main = 1.2, font.main = 1)

pie(table(dat$topo), main = "Topography",radius = 0.9, cex = 0.8, cex.main = 1.2, font.main = 1)
library(lattice)
library(latticeExtra)

dat2001 <- subset(dat, year == 2001)

topo_cols <- adjustcolor(c("#C44E52","#4E9A06","#4C72B0","#DD8452"), alpha.f = 0.7)

legend_labels <- c("W = West slope", "HT = Hilltop", "E = East slope", "LO = Low East")
xyplot(lat ~ long, data = dat2001, groups = topo, pch = 16, col = topo_cols,
  key = list(
    space = "right", title = "Topography",
    points = list(col = topo_cols, cex = 1.2, pch = 15),
```

```r
    text = list(legend_labels)
  )
)
redblue <- colorRampPalette(c("firebrick", "lightgray", "#375997"))
levelplot(yield ~ long*lat|factor(year), data=dat,
          xlab="Longitude", ylab="Latitude",
          scales=list(alternating=FALSE),
          prepanel = prepanel.default.xyplot,
          panel = panel.levelplot.points,
          type = c("p", "g"), aspect = "iso", col.regions=redblue, cex = 0.4)
means <- replicate(10000,mean(sample(dat2001$yield, size = 30, replace = TRUE)))
hist(means, main = "", xlab = "mean yield (quintals/ha)")
dat2001hills <- subset(dat2001, topo == "HT")
qqnorm(dat2001hills$yield,main = "", pch = 16, cex = 0.4, col = rgb(1, 0, 0, 0.5)); qqline(dat2001hills$
ttest_result <- t.test(dat2001hills$yield, mu = 75.2, conf.level = 0.99)
results_table <- tibble(
  "p-value" =  format(ttest_result$p.value, scientific = TRUE, digits = 2),
  df = ttest_result$parameter,
  "99% CI" = paste(round(ttest_result$conf.int[1], 2), "-",
                   round(ttest_result$conf.int[2], 2))
)

knitr::kable(results_table,
             row.names = FALSE,
             align = c("l", "l", "l"))
sample_data %>%
  group_by(cut(x, breaks = 3, labels = c("Low", "Medium", "High"))) %>%
  summarize(
    Count = n(),
    Mean_Y = mean(y),
    SD_Y = sd(y),
    Min_Y = min(y),
    Max_Y = max(y),
    .groups = "drop"
  ) %>%
  kable(
    digits = 2,
    col.names = c("X Category", "Count", "Mean Y", "SD Y", "Min Y", "Max Y"),
    align = "lccccc"
  ) %>%
  kable_classic() %>%
  add_footnote(
    label = c("Data grouped by tertiles of x values."),
    notation = "none",
    threeparttable = TRUE
  )
ggplot(sample_data, aes(x = x, y = y)) +
  geom_point(alpha = 0.7, color = "steelblue") +
  geom_smooth(method = "lm", se = TRUE, color = "darkred") +
  theme_minimal() +
  labs(
    x = "Predictor (x)",
    y = "Response (y)"
```

```r
  )
ggplot(sample_data, aes(x = x)) +
  geom_histogram(
    binwidth = 2,
    fill = "steelblue",
    color = "black",
    alpha = 0.7
  ) +
  theme_minimal() +
  labs(
    x = "Predictor (x)",
    y = "Frequency",
    alt = "Histogram showing the distribution of x values, which appears approximately normal and center
  )
# Reshape data for plotting
plot_data <- sample_data %>%
  select(x, y) %>%
  pivot_longer(cols = everything(), names_to = "variable", values_to = "value")

ggplot(plot_data, aes(x = variable, y = value, fill = variable)) +
  geom_boxplot(alpha = 0.7) +
  theme_minimal() +
  scale_fill_brewer(palette = "Set2") +
  labs(
    x = "Variable",
    y = "Value",
    fill = "Variable"
  )
```