Final Project Report

Ananya D, Ekansh B, Hannah Wu

Analysis of Cotton and Cottonseed Prices (1910-1938)

Introduction

The economic significance of cotton and cottonseed during the early 20th century was profound, impacting global trade, agriculture, and industrial processes. This analysis focuses on examining trends and relationships in cotton and cottonseed prices between 1910 and 1938, a period marked by economic upheaval and agricultural innovation.

The data set `Merged_Cotton_Cottonseed_Prices.csv` was created by cleaning and filtering two primary data sources: `Cotton_Wholesale_Prices_1910_1938.csv` and `Cottonseed_Prices_1910_1938.csv`. The goal of this project is to uncover insights into price trends, identify key patterns, and provide a foundation for understanding the economic dynamics of this period.

Literature Review

Studies on historical commodity pricing often underscore the importance of agricultural goods in shaping economic policies and societal outcomes. Cotton, often referred to as "white gold," played a pivotal role in the economies of the United States and other countries during the early 20th century. Similarly, cottonseed, a byproduct of cotton production, was increasingly valued for its use in oil production and livestock feed.

Previous research has highlighted: - The influence of economic events such as the Great Depression on commodity prices. - The relationship between agricultural yields and market fluctuations. - The evolution of byproduct markets, such as cottonseed oil, which gained prominence during this era.

Methodology

Data Sources

The dataset was constructed by integrating and cleaning the following:

1. **Cotton Wholesale Prices (1910–1938)**: Provided annual wholesale price data for raw cotton.
2. **Cottonseed Prices (1870–1945)**: Contained corresponding annual prices for cottonseed.

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union


Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

    group_rows

Warning: NAs introduced by coercion
```

Cleaning and Filtering Process

The original data sets were pre-processed to: - Remove missing or incomplete records. - Align the time frames and ensure consistency in units of measurement. - Retain only relevant columns (e.g., year, price, region).

Cleaned Data:

```
  Year Month Cottonseed_Price Cotton_Wholesale_Price
1 1910     9            26.23                   14.0
2 1910    10            26.86                   14.5
3 1910    11            25.36                   14.8
4 1910    12            25.65                   15.1
5 1911     1            26.35                   14.9
6 1911     2            25.61                   14.3
```

Analytical Tools

- **Data visualization**: Boxplots, histograms, and line graphs were used to identify trends and outliers.
- **Summary statistics**: Measures of central tendency and dispersion were calculated for both cotton and cottonseed prices.
- **Correlation analysis**: Examined the relationship between cotton and cottonseed prices over time.

Data Exploration

Visualizing Price Trends

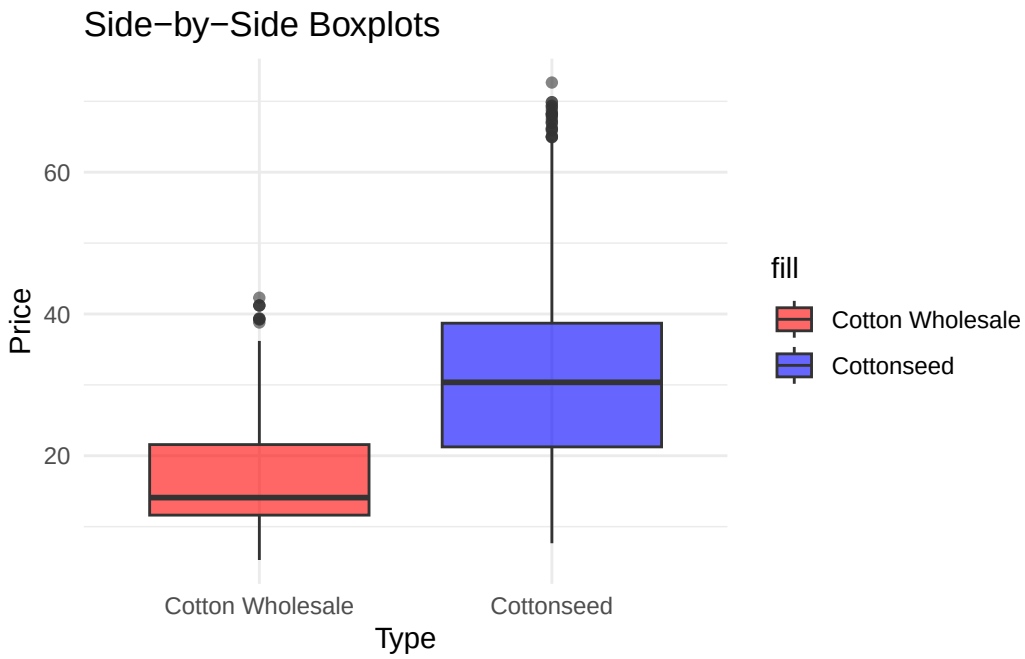
The merged dataset revealed distinct price trends for cotton and cottonseed.

Summary Statistics A summary of cotton and cottonseed prices is presented below:

Table 1: Summary Statistics for Cotton and Cottonseed Prices (1910–1938)

Year	Month	Cottonseed_Price	Cotton_Wholesale_Price
Min. :1910	Min. : 1.000	Min. : 7.66	Min. : 5.30
1st Qu.:1917	1st Qu.: 3.250	1st Qu.:21.25	1st Qu.:11.62
Median :1924	Median : 6.000	Median :30.36	Median :14.10
Mean :1924	Mean : 6.494	Mean :32.37	Mean :17.27
3rd Qu.:1931	3rd Qu.: 9.750	3rd Qu.:38.70	3rd Qu.:21.57
Max. :1938	Max. :12.000	Max. :72.65	Max. :42.30

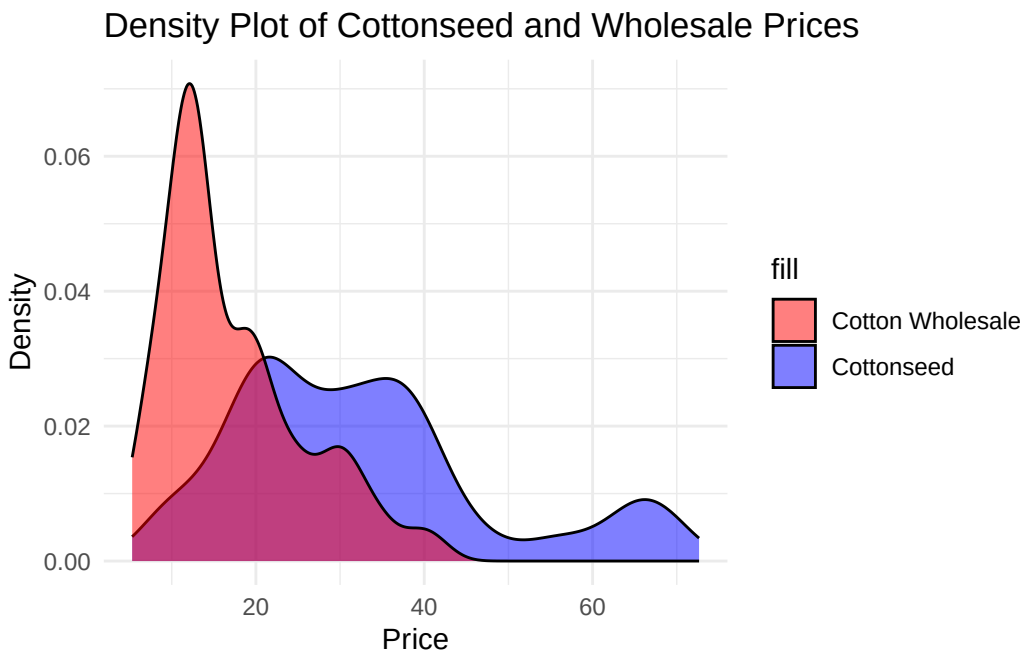
Below is an example of a boxplot comparing annual price distributions:



Why this plot is important:

- Boxplot show the range, interquartile range (IQR), and any outliers.
- They are ideal for identifying years with unusual price spikes or dips.

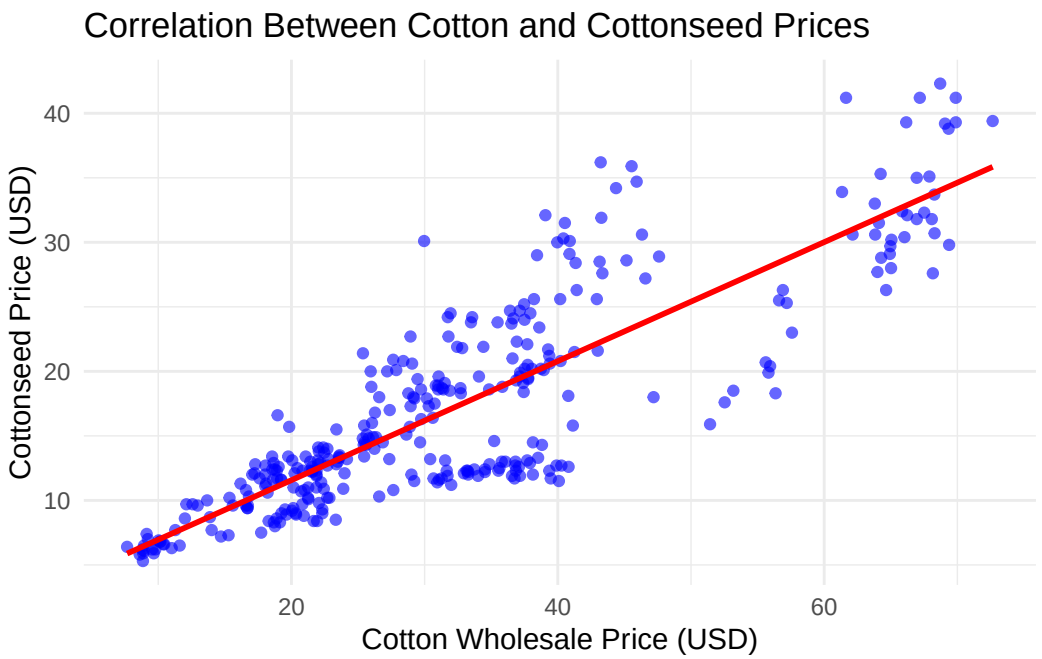
Below is a density plot…



Scatter Plot: Price Correlation

A scatter plot was generated to examine the relationship between cotton and cottonseed prices.

```
`geom_smooth()` using formula = 'y ~ x'
```



Why this plot is important:

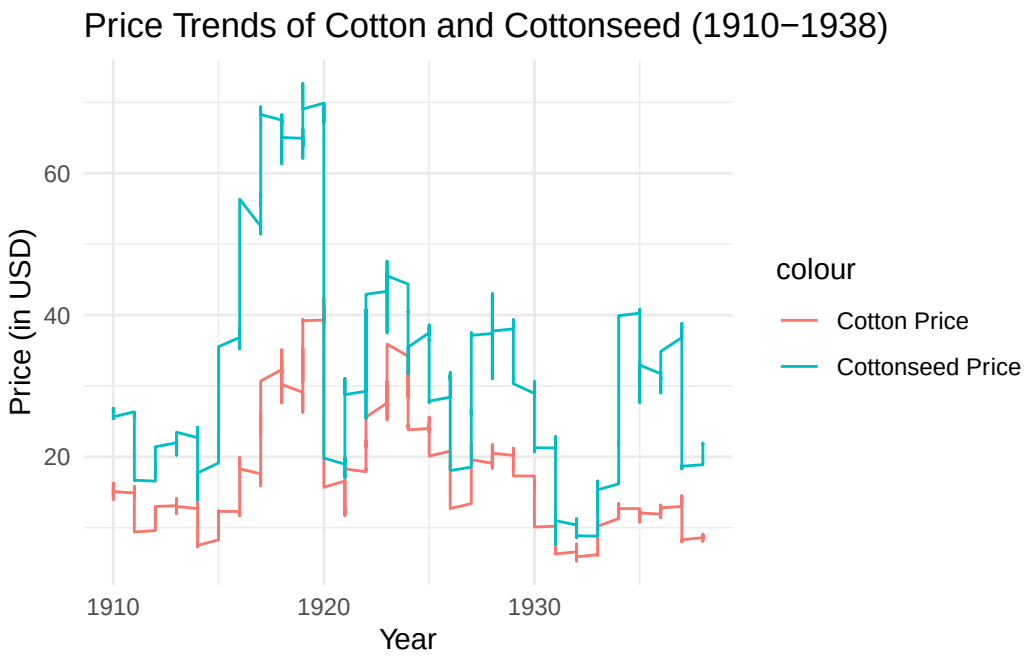
- Scatter plots reveal the strength and direction of relationships between two variables.
- They indicate whether changes in cotton prices influence cottonseed prices.

Price Trends Over Time

A **Line graph** plotted for both cotton and cottonseed prices to track price trends from 1910 to 1938.

```
Warning: No shared levels found between `names(values)` of the manual scale and the
data's fill values.
```

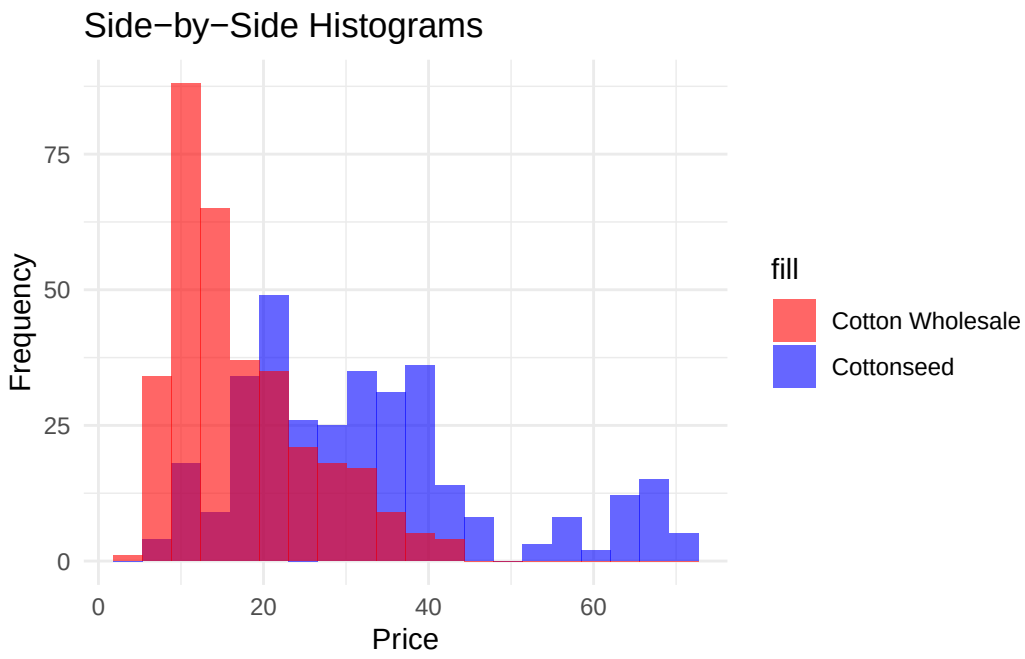
```
Warning: Removed 13 rows containing missing values or values outside the scale range
(`geom_line()`).
```



Why this plot is important: The Line graphs provides an intuitive way to visualize trends, highlighting periods of price volatility or stability. They help us compare cotton and cottonseed prices over time.

Histogram...

A **Histogram** was created to visualize the frequency distribution of prices for cotton and cottonseed.



Why this plot is important: The Histogram reveals the shape of the price distribution (e.g., normal, skewed). It helps identify common price ranges and outlier frequencies.

Results and Discussion

- **Observed Patterns Price Variability:** Cottonseed prices demonstrated greater variability compared to cotton prices, likely reflecting its emerging market status.
- **Trends Over Time:** Both commodities exhibited significant price fluctuations during the Great Depression, underscoring the economic impact of the era.
- **Correlation:** A positive correlation was observed between cotton and cottonseed prices, suggesting interdependence in their market dynamics.
- **Outliers:** Certain years, such as 1929 and 1933, showed extreme deviations in prices, possibly linked to economic policies or natural events affecting production (like the Great Depression).

Conclusion

The analysis of cotton and cottonseed prices from 1910 to 1938 highlights their intertwined economic importance. Key findings include:

The variability and trends in pricing reflect broader economic conditions of the era.

The relationship between the two commodities indicates shared market drivers.

Further analysis could explore regional differences or integrate external factors such as production volumes and policy changes.

References Source datasets:

- Cotton_Wholesale_Prices_1910_1938.csv.
- Cottonseed_Prices_1910_1938.csv.

"Agricultural crises and the international transmission of the Great Depression on JSTOR." www.jstor.org. JSTOR, www.jstor.org/stable/2698023.

- Additional references to be determined.

##Code Appendix