Festive Spending Analysis: An Econometric Exploration

AUTHOR
Vaishnawi Sharma

PUBLISHED
February 26, 2025

Introduction

Hey everyone! This is a fun little project where I'm diving into the world of Economics and Econometrics, and using R to bring it all to life. I'm basically trying to see how festive seasons affect how much people spend in different regions. Think of it like exploring shopping habits during Diwali, Christmas, and other celebrations! My professor suggested I play around with some cool graphs and see what insights I can find. So, let's get started!

Loading a Few Packages

First things first, we need to bring in some helpful tools from the R world. These are called "packages."

```
library(ggplot2) # This package is like the artist's toolkit for making beautiful graphs. We'll us
library(dplyr)   # This one is a data manipulation wizard! It helps us wrangle and organize our da
```

```
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
```

```
library(forcats) # This helps us reorder factors based on frequency.
```

Simulated Data

Since we don't have real data (yet!), we'll create some fake data to work with. This is called "simulated data." It's like a practice run before we work with the real thing.

```
# Let's create a data frame with the required data.
set.seed(123)
festive_spending <- data.frame(
  Region = rep(c("North", "South", "East", "West", "Central"), each = 4),
  Festival = rep(c("Diwali", "Christmas", "Eid", "Holi"), times = 5),
  AvgSpend = round(runif(20, min = 5000, max = 30000), 2),
```

```
  NumTransactions = round(runif(20, min = 10, max = 100))
)

# Let's take a peek at our newly created data frame!
head(festive_spending)
```

```
  Region  Festival AvgSpend NumTransactions
1  North    Diwali 12189.44              90
2  North Christmas 24707.63              72
3  North       Eid 15224.42              68
4  North      Holi 27075.44              99
5  South    Diwali 28511.68              69
6  South Christmas  6138.91              74
```

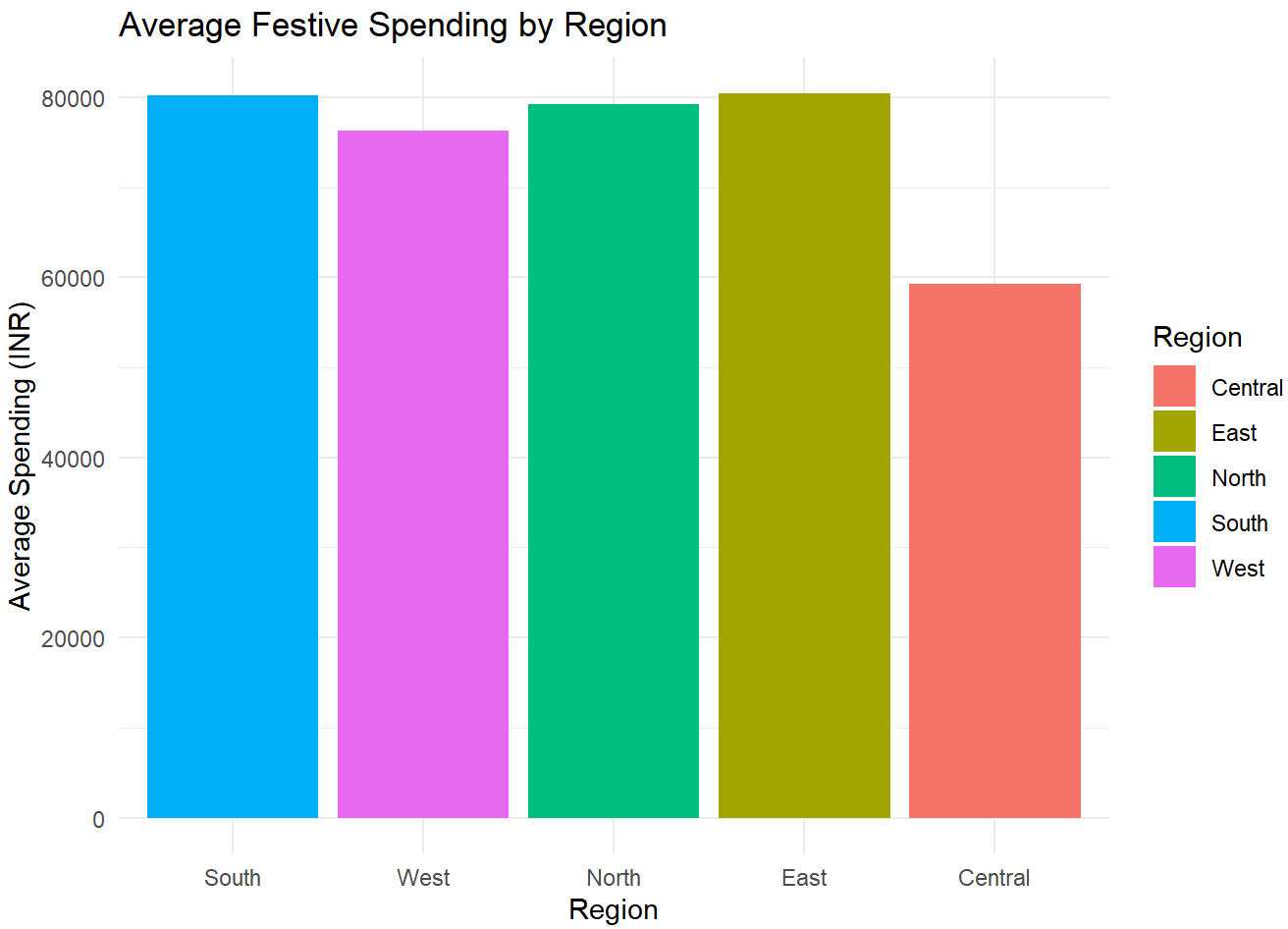
Visualizations

Now for the fun part – creating some visuals!

(a) Bar plot: AvgSpend by Region

We'll start with a bar plot to see how average spending varies across different regions.

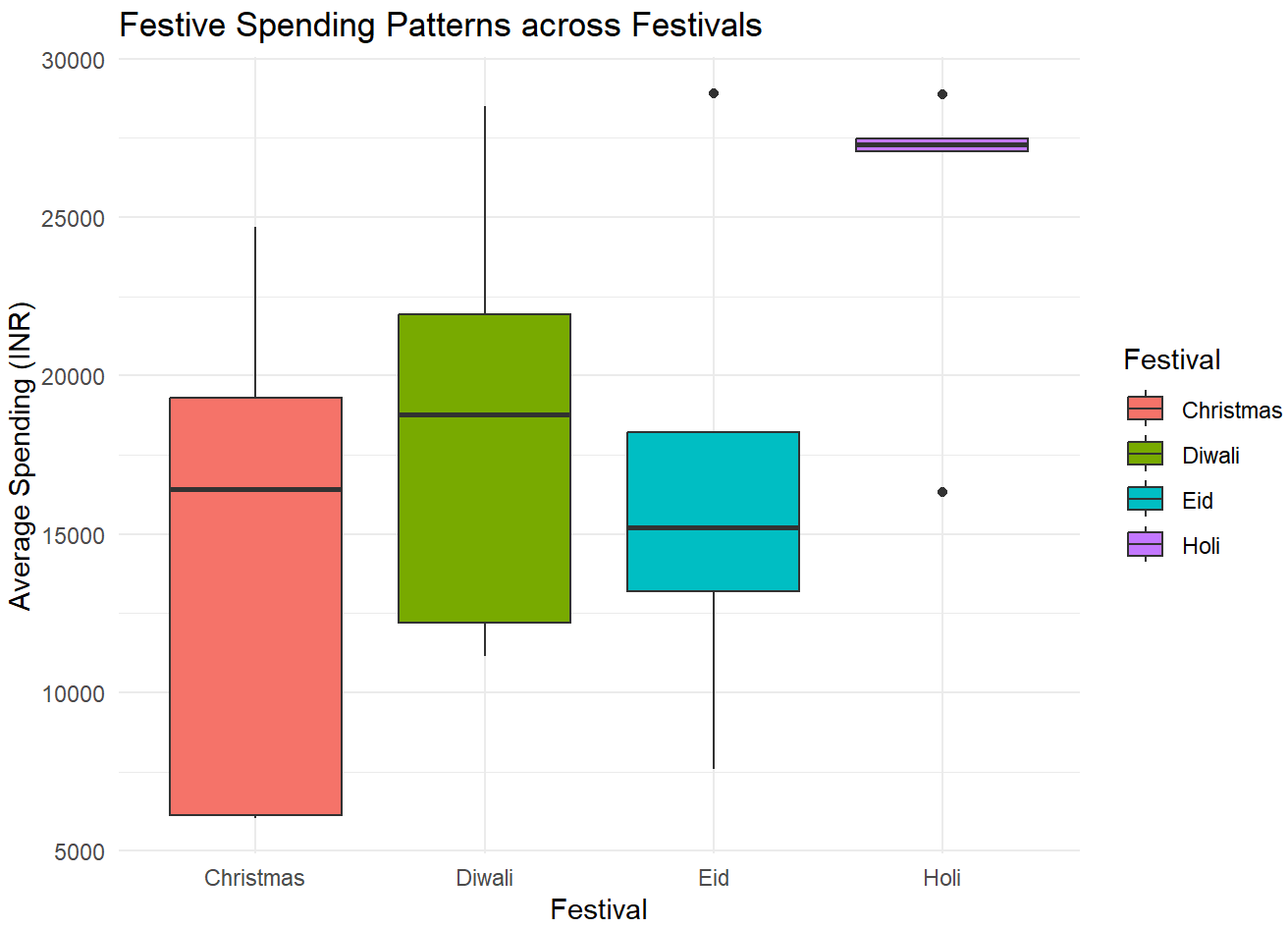
```
# Bar plot with fct_reorder.
ggplot(data = festive_spending,
       mapping = aes(x = fct_reorder(Region, AvgSpend, .desc = TRUE),
                     y = AvgSpend, fill = Region)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Festive Spending by Region",
       x = "Region",
       y = "Average Spending (INR)") +
  theme_minimal()
```



(b) Boxplot: AvgSpend by Festival

Next, a boxplot to see how spending patterns change across different festivals.

```
ggplot(data = festive_spending,
       mapping = aes(x = Festival, y = AvgSpend, fill = Festival)) +
  geom_boxplot() +
  labs(title = "Festive Spending Patterns across Festivals",
       x = "Festival",
       y = "Average Spending (INR)") +
  theme_minimal()
```

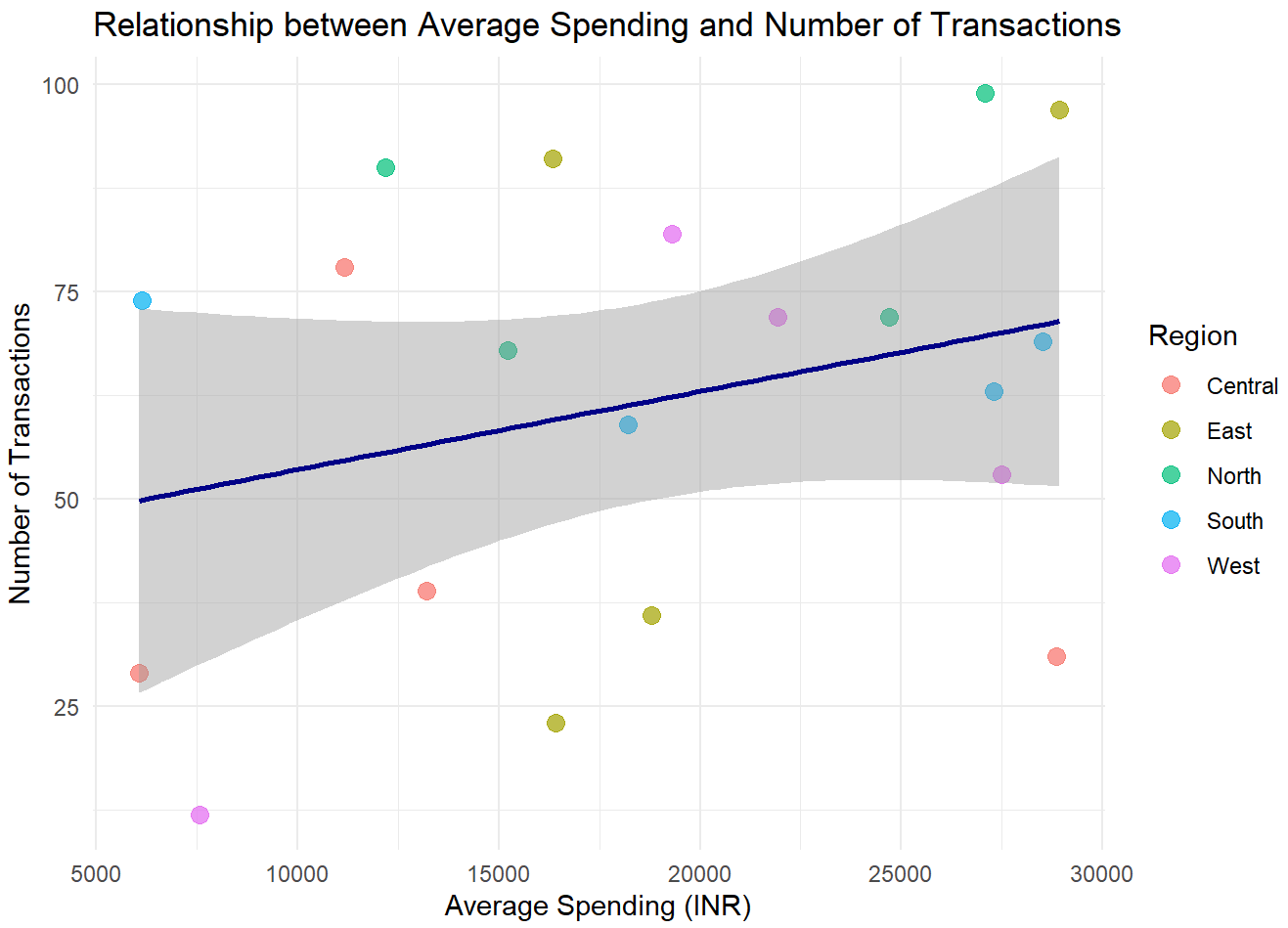


(c) Scatter plot: AvgSpend vs NumTransactions

Initially, I tried making a scatter plot with multiple aesthetic mappings, but it became cluttered and hard to interpret. A conventional scatter plot usually works best when comparing two continuous variables. Therefore, I simplified the plot to focus solely on the relationship between average spending and the number of transactions. To further enhance understanding, I added a trendline using `geom_smooth()`. By default, `geom_smooth()` shows a 95% confidence interval, indicating the range within which we expect the true trend to fall.

```
# Improved Scatter Plot for Continuous Variables
ggplot(festive_spending, aes(x = AvgSpend, y = NumTransactions)) +
  geom_point(aes(color = Region), size = 3, alpha = 0.7) +
  geom_smooth(method = "lm", color = "darkblue", se = TRUE) +
labs(title = "Relationship between Average Spending and Number of Transactions",
       x = "Average Spending (INR)",
       y = "Number of Transactions") +
  theme_minimal()
```

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

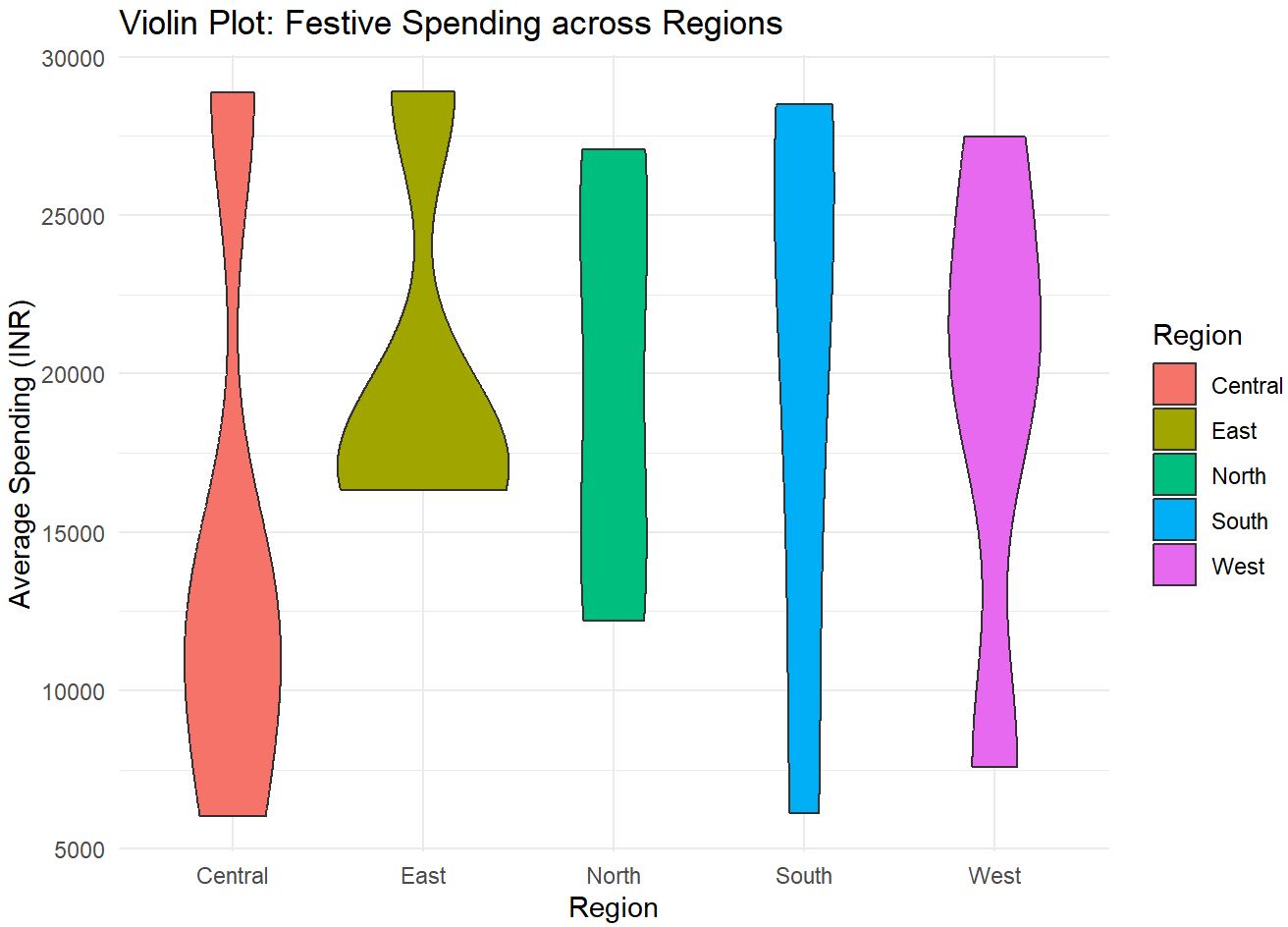


This approach simplifies the visualization, allowing us to see whether there's a positive, negative, or no correlation between spending and transactions.

Alternative Plot: Violin plot for AvgSpend by Region

Just for fun, let's try a different type of plot – a violin plot! It's similar to a boxplot but gives a better idea of the distribution of the data.

```
ggplot(data = festive_spending,
       mapping = aes(x = Region, y = AvgSpend, fill = Region)) +
  geom_violin() +
  labs(title = "Violin Plot: Festive Spending across Regions",
       x = "Region",
       y = "Average Spending (INR)") +
  theme_minimal()
```



Insights

Now, let's look at what the graphs tell us!

1. **Region and Spending:** It seems like people in the North and West tend to spend more during festivals. Maybe they have bigger celebrations or more disposable income?
2. **Festival and Spending:** Diwali and Christmas seem to be the biggest spending festivals. This probably reflects the cultural importance and lots of shopping that happens around these times.
3. **Transactions:** Higher spending often coincides with more transactions, but some regions show high spending with fewer transactions, suggesting luxury purchases.

Conclusion

This was a short and sweet project where I got to play around with R and explore how festive seasons influence spending. It was a great way to combine my economics and econometrics knowledge with some practical data visualization. I learned a lot, and hopefully, you did too!