Analyzing the Growth of Warehouses in Southern California

Yatziri I. Carmona Ochoa

2025-11-20

```
library(readxl)
library(dplyr)
```

```
##
## 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(ggplot2)
library(janitor)
```

```
## Warning: package 'janitor' was built under R version 4.5.2
```

```
##
## Attaching package: 'janitor'
```

```
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

```
# Read data
df <- read_excel("Warehouse_List.xlsx")
```

```
## New names:
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
```

```r
# Drop header row & rename category column
df_clean <- df[-1, ]
names(df_clean)[1] <- "category"

# Extract useful numeric fields
df_clean <- df_clean |>
mutate(
year_built = as.numeric(`...4`),
building_sq_ft = as.numeric(`...6`)
) |>
filter(!is.na(year_built))
```

```r
#Count warehouses by category
df_clean |>
group_by(category) |>
summarise(total = n(), .groups = "drop")
```

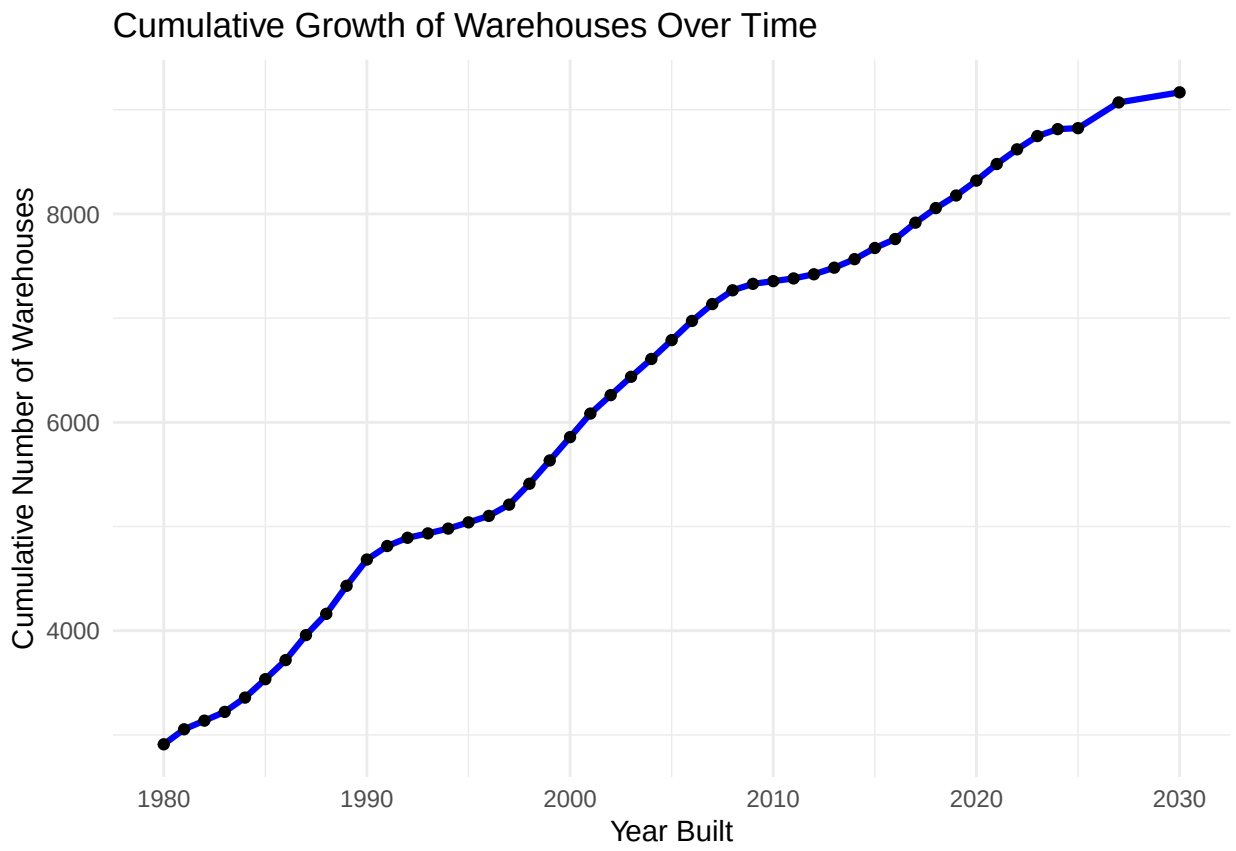
```
## # A tibble: 3 x 2
##   category    total
##   <chr>       <int>
## 1 Approved      243
## 2 CEQA Review   101
## 3 Existing     8823
```

Here, I'm analyzing the cumulative growth of all warehouse categories (Existing, Approved, and CEQA Review)

```r
#Cumulative growth of all warehouse categories
warehouse_counts <- df_clean |>
group_by(year_built) |>
summarise(count = n(), .groups = "drop") |>
arrange(year_built) |>
mutate(cumulative = cumsum(count))
```

```r
#Plot of cumulative warehouse growth for all categories
ggplot(warehouse_counts, aes(year_built, cumulative)) +
geom_line(color = "blue", size = 1.1) +
geom_point(color = "black") +
labs(
title = "Cumulative Growth of Warehouses Over Time",
x = "Year Built",
y = "Cumulative Number of Warehouses"
) +
theme_minimal()
```

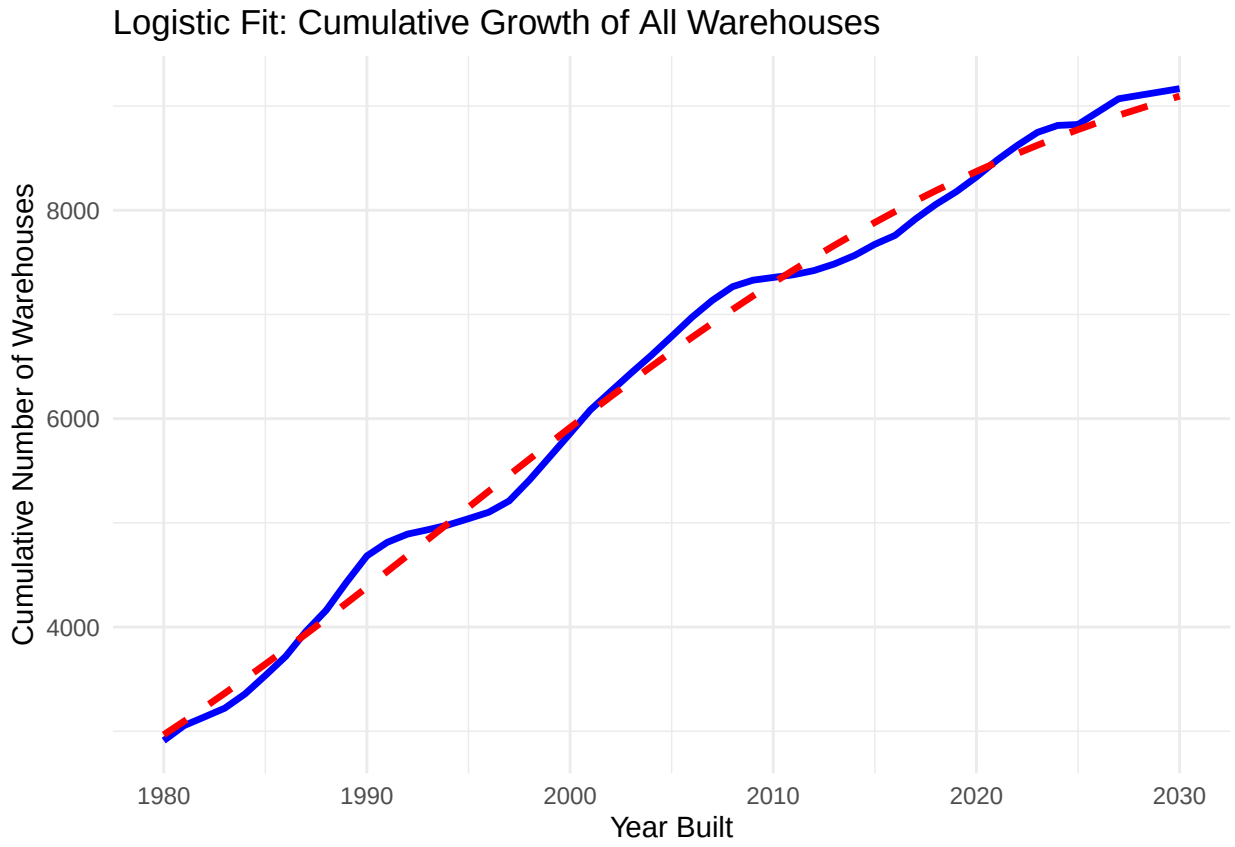
```
## Warning: Using ‘size‘ aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use ‘linewidth‘ instead.
## This warning is displayed once every 8 hours.
## Call ‘lifecycle::last_lifecycle_warnings()‘ to see where this warning was
## generated.
```




```
#Create a logistic growth for cumulative warehouse growth for all categories

warehouse_counts <- warehouse_counts |>
mutate(t = year_built - min(year_built))

logistic_all <- nls(
cumulative ~ K / (1 + A * exp(-r * t)),
data = warehouse_counts,
start = list(
K = max(warehouse_counts$cumulative) * 1.2,
A = 1,
r = 0.1
)
)

warehouse_counts <- warehouse_counts |>
mutate(logistic_pred = predict(logistic_all))
```

```
#Overlay logistic curve with plot for cumulative warehouse growth across all categories

ggplot(warehouse_counts, aes(year_built)) +
geom_line(aes(y = cumulative), color = "blue", size = 1.2) +
geom_line(aes(y = logistic_pred), color = "red", linetype = "dashed", size = 1.2) +
labs(
title = "Logistic Fit: Cumulative Growth of All Warehouses",
y = "Cumulative Number of Warehouses",
```

```
x = "Year Built"
) +
theme_minimal()
```

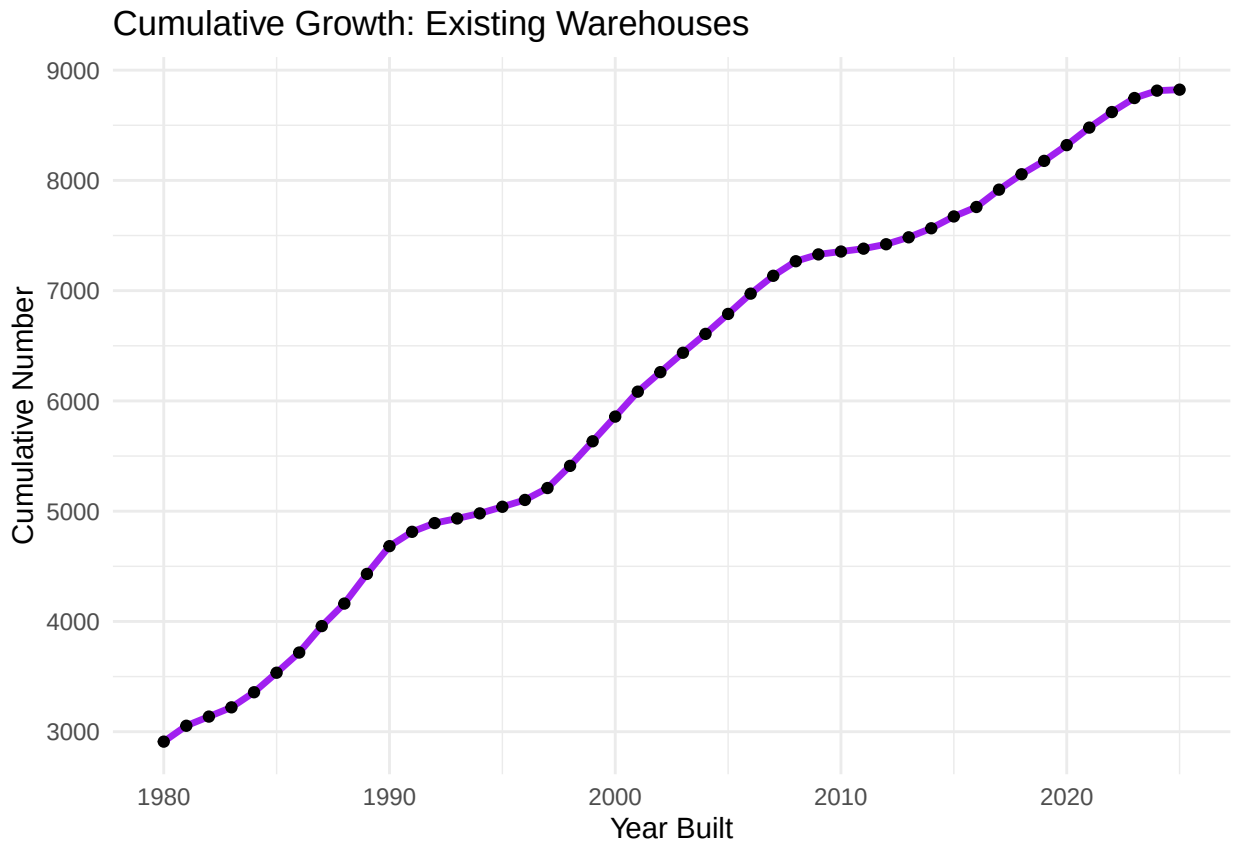


Now, let's analyze only existing warehouses.

```
#Count existing warehouses
df_existing <- df_clean |>
filter(category == "Existing")

existing_counts <- df_existing |>
group_by(year_built) |>
summarise(count = n(), .groups = "drop") |>
arrange(year_built) |>
mutate(cumulative = cumsum(count))
```

```
#Plot cumulative growth of existing warehouses

ggplot(existing_counts, aes(year_built, cumulative)) +
geom_line(color = "purple", size = 1.2) +
geom_point(color = "black") +
theme_minimal() +
labs(
title = "Cumulative Growth: Existing Warehouses",
x = "Year Built",
```

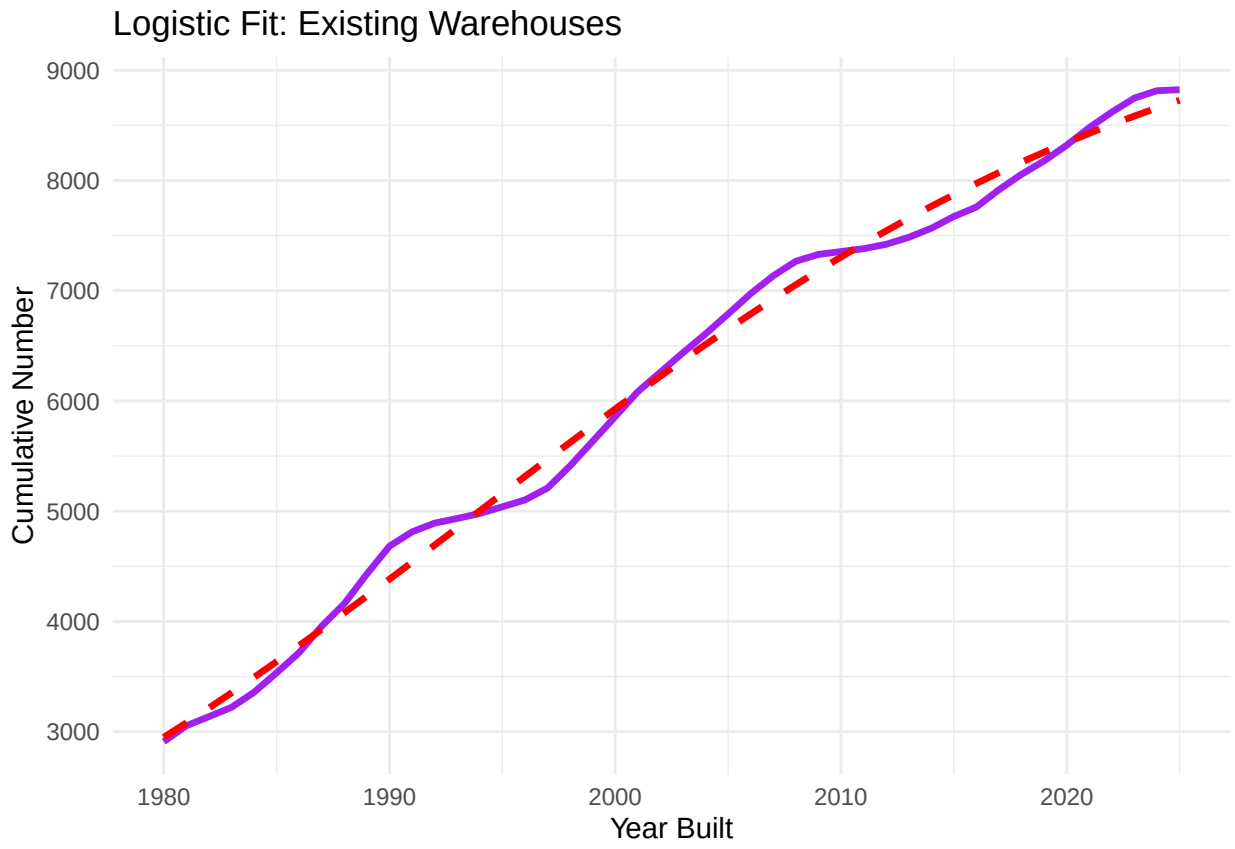
```
y = "Cumulative Number"
)
```




```
#Create logistic curve

existing_counts <- existing_counts |>
mutate(t = year_built - min(year_built))

logistic_existing <- nls(
cumulative ~ K / (1 + A * exp(-r * t)),
data = existing_counts,
start = list(
K = max(existing_counts$cumulative) * 1.2,
A = 1,
r = 0.1
)
)

existing_counts <- existing_counts |>
mutate(logistic_pred = predict(logistic_existing))
```

```
#Overlay logistic curve with plot for cumulative growth for existing Warehouses

ggplot(existing_counts, aes(year_built)) +
geom_line(aes(y = cumulative), color = "purple", size = 1.2) +
```

```
geom_line(aes(y = logistic_pred), color = "red", linetype = "dashed", size = 1.2) +
theme_minimal() +
labs(
title = "Logistic Fit: Existing Warehouses",
x = "Year Built",
y = "Cumulative Number"
)
```

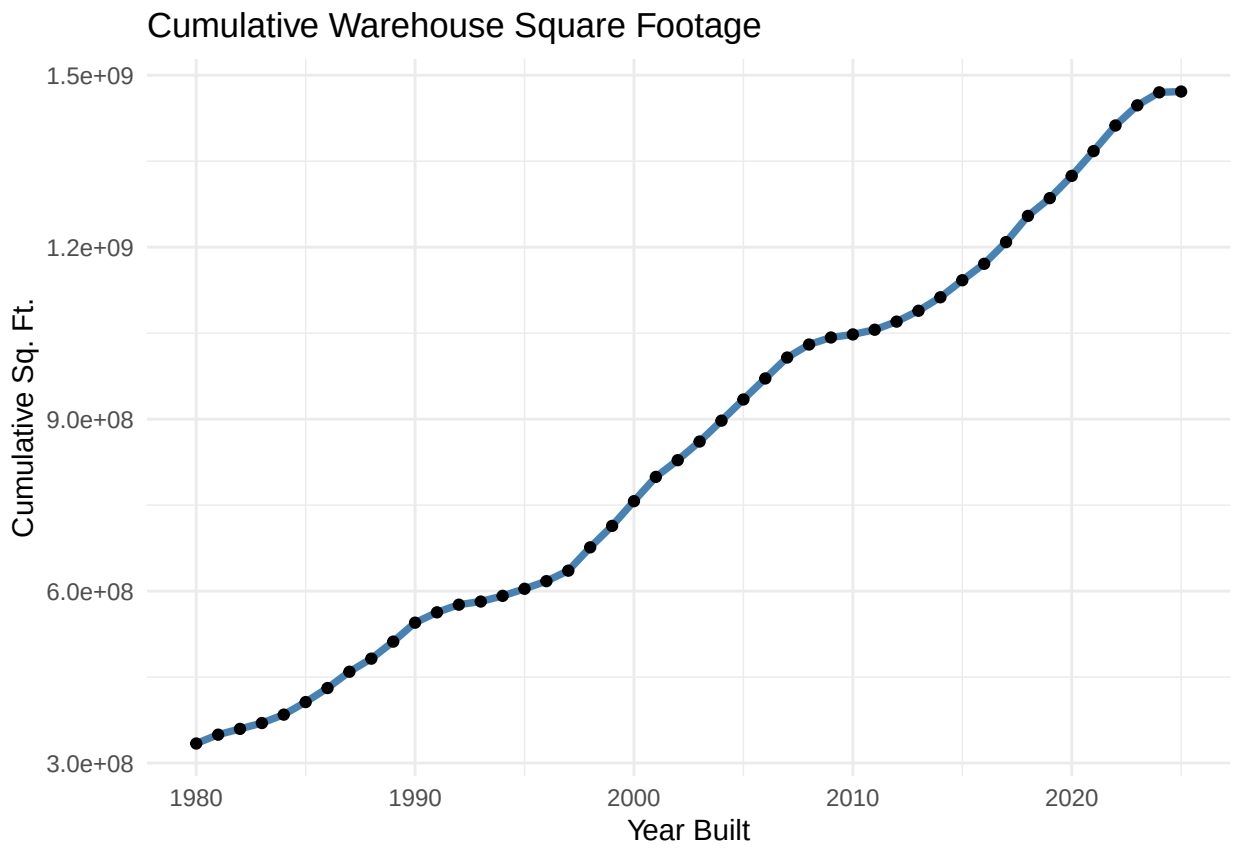


Now, let's analyze the square Footage of the existing warehouse buildings

```
#Count square footage for existing warehouse buildings
sqft_yearly <- df_existing |>
group_by(year_built) |>
summarise(total_sqft = sum(building_sq_ft, na.rm = TRUE), .groups = "drop") |>
arrange(year_built) |>
mutate(cumulative_sqft = cumsum(total_sqft))
```

```
#Plot cumulative square footage for existing warehouse buildings

ggplot(sqft_yearly, aes(year_built, cumulative_sqft)) +
geom_line(color = "steelblue", size = 1.3) +
geom_point(color = "black") +
theme_minimal() +
labs(
title = "Cumulative Warehouse Square Footage",
```

```
x = "Year Built",
y = "Cumulative Sq. Ft."
)
```




```
#Create a logistic Curve

sqft_yearly <- sqft_yearly |>
mutate(t = year_built - min(year_built))

sqft_fit <- nls(
cumulative_sqft ~ K / (1 + A * exp(-r * t)),
data = sqft_yearly,
start = list(
K = max(sqft_yearly$cumulative_sqft) * 1.2,
A = 1,
r = 0.05
)
)

sqft_yearly <- sqft_yearly |>
mutate(logistic_pred = predict(sqft_fit))
```

```
#Plot for Logistic Curve Overlay
ggplot(sqft_yearly, aes(year_built)) +
geom_line(aes(y = cumulative_sqft), color = "steelblue", size = 1.2) +
```

```
geom_line(aes(y = logistic_pred), color = "red", size = 1.2, linetype = "dashed") +
theme_minimal() +
labs(
title = "Logistic Fit: Cumulative Square Footage",
x = "Year Built",
y = "Cumulative Sq. Ft."
)
```