Final Research Report

Thomas Soupionis

2025-05-16

0. Load R Packages

```
# tidyverse for data wrangling and ggplot2 graphics, lubridate for date parsing
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## Warning: package 'forcats' was built under R version 4.3.3
```

```
## Warning: package 'lubridate' was built under R version 4.3.2
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.0
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(lubridate)
library(ggplot2)
```

1. Research Question and Variable Definitions

Research Question: What are the distributions of 2-, 5-, and 10-year returns for the Dow Jones Industrial Average, S&P 500, and NASDAQ Composite, and how does holding period affect return consistency? Additionally, how do these outcomes differ when we limit the analysis to periods beginning at the S&P 500 inception (March 4, 1957)?

Dependent Variable: - **Return_pct:** Numeric variable representing the percent return over a specified holding period (2, 5, or 10 years).

Independent Variables: - **Horizon:** Categorical variable indicating holding period with three levels: "2 Years", "5 Years", and "10 Years". - **Index:** Categorical variable indicating the market index with three levels: "Dow Jones", "S&P 500", and "NASDAQ".

2. Data Import and Preparation (Full Data)

Data Source: ^SPX, ^NDQ, and ^DJI on Stooq.com where all columns are deleted except Date and Close

```
# Read daily closing data and add an "Index" identifier
# ymd() and mdy() convert character dates to Date objects

dow_data    <- read_csv("../data/official_dowjones_data.csv", col_types = cols(.default = col_guess(), I
  mutate(
    # parse both YYYY-MM-DD and M/D/YYYY date formats
    Date = parse_date_time(Date, orders = c("Y-m-d", "m/d/Y")),
    Index = "Dow Jones"
  )

spx_data    <- read_csv("../data/official_spx_data.csv") %>%
  mutate(
    Date = mdy(Date),
    Index = "S&P 500")
```

```
## Rows: 17168 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Date
## dbl (1): Close
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
nasdaq_data <- read_csv("../data/official_nasdaq_data.csv") %>%
  mutate(Date = mdy(Date),
         Index = "NASDAQ")
```

```
## Rows: 13686 Columns: 2
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): Date
## dbl (1): Close
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Function to compute percent return over 2-, 5-, and 10-year lags
# 504 trading days is about 2 years, 1260 is about 5 years, 2520 is about 10 years
calculate_returns <- function(df) {
  df %>%
    arrange(Date) %>%
    mutate(
      Return_2yr  = (Close / lag(Close, 504)  - 1) * 100,
      Return_5yr  = (Close / lag(Close, 1260) - 1) * 100,
      Return_10yr = (Close / lag(Close, 2520) - 1) * 100
    )
}
```

```r
# Apply the function to each index series
dow_ret    <- calculate_returns(dow_data)
spx_ret    <- calculate_returns(spx_data)
nasdaq_ret <- calculate_returns(nasdaq_data)

# Combine all into one long data frame for plotting
all_returns <- bind_rows(dow_ret, spx_ret, nasdaq_ret) %>%
  select(Date, Index, starts_with("Return")) %>%
  pivot_longer(
    cols      = starts_with("Return"),
    names_to  = "Horizon",
    values_to = "Return_pct"
  ) %>%
  mutate(
    Horizon = recode(Horizon,
      Return_2yr  = "2 Years",
      Return_5yr  = "5 Years",
      Return_10yr = "10 Years"
    ),
    # set factor levels for correct order
    Horizon = factor(Horizon, levels = c("2 Years", "5 Years", "10 Years"))
  ) %>%
  filter(!is.na(Return_pct))

head(all_returns)
```

```
## # A tibble: 6 x 4
##   Date                Index     Horizon Return_pct
##   <dttm>              <chr>     <fct>        <dbl>
## 1 1898-05-25 00:00:00 Dow Jones 2 Years       27.8
## 2 1898-05-26 00:00:00 Dow Jones 2 Years       29.1
## 3 1898-05-27 00:00:00 Dow Jones 2 Years       28.2
## 4 1898-05-31 00:00:00 Dow Jones 2 Years       29.9
## 5 1898-06-01 00:00:00 Dow Jones 2 Years       32.0
## 6 1898-06-02 00:00:00 Dow Jones 2 Years       34.2
```

3. Data Import and Preparation (Post-SPX Inception)

Data Import and Preparation (Post-SPX Inception)

```r
# Restrict Dow to post S&P inception date
spx_start <- min(spx_data$Date)
dow_data_f <- dow_data %>% filter(Date >= spx_start)
# Recompute returns for filtered Dow
dow_ret_f <- calculate_returns(dow_data_f)

# Combine and reshape (filtered)
all_returns_filt <- bind_rows(dow_ret_f, spx_ret, nasdaq_ret) %>%
  select(Date, Index, starts_with("Return")) %>%
  pivot_longer(cols = starts_with("Return"),
               names_to = "Horizon", values_to = "Return_pct") %>%
```

```
  mutate(
    Horizon = recode(Horizon,
      Return_2yr  = "2 Years",
      Return_5yr  = "5 Years",
      Return_10yr = "10 Years"),
    Horizon = factor(Horizon, levels = c("2 Years", "5 Years", "10 Years"))
  ) %>%
  filter(!is.na(Return_pct))

head(all_returns_filt)
```

```
## # A tibble: 6 x 4
##   Date                Index     Horizon Return_pct
##   <dttm>              <chr>     <fct>        <dbl>
## 1 1959-03-04 00:00:00 Dow Jones 2 Years       29.8
## 2 1959-03-05 00:00:00 Dow Jones 2 Years       29.4
## 3 1959-03-06 00:00:00 Dow Jones 2 Years       28.3
## 4 1959-03-09 00:00:00 Dow Jones 2 Years       28.6
## 5 1959-03-10 00:00:00 Dow Jones 2 Years       29.6
## 6 1959-03-11 00:00:00 Dow Jones 2 Years       30.2
```

4. Summary Statistics on Full vs Filtered

```
# Calculate mean, median, standard deviation, and positive/negative return percentages per horizon
get_summary <- function(data, label) {
  data %>% group_by(Horizon) %>%
    summarize(
      Scenario      = label,
      Mean_Return   = mean(Return_pct),
      Median_Return = median(Return_pct),
      SD_Return     = sd(Return_pct),
      Pct_Positive  = mean(Return_pct > 0) * 100,
      Pct_Negative  = mean(Return_pct < 0) * 100
    )
}

sum_full <- get_summary(all_returns, "Full Data")
sum_filt <- get_summary(all_returns_filt, "Post-1957")

# Combine and display
results <- bind_rows(sum_full, sum_filt)
knitr::kable(results, digits = 2,
             caption = "Summary Statistics for Full vs Post-SPX Inception Data")
```

Table 1: Summary Statistics for Full vs Post-SPX Inception Data

Horizon	Scenario	Mean_Return	Median_Return	SD_Return	Pct_Positive	Pct_Negative
2 Years	Full Data	18.33	17.24	29.01	76.34	23.63
5 Years	Full Data	50.81	45.64	58.33	80.70	19.29

Horizon	Scenario	Mean_Return	Median_Return	SD_Return	Pct_Positive	Pct_Negative
10 Years	Full Data	119.61	99.64	114.13	89.17	10.82
2 Years	Post-1957	19.56	18.20	26.80	80.23	19.75
5 Years	Post-1957	56.19	50.87	57.85	83.29	16.70
10 Years	Post-1957	139.82	126.97	121.50	91.79	8.21

4.1 Interpreting Summary Statistics

- **SD >= mean**: Returns are volatile relative to their average, indicating higher risk.
- **SD <= mean**: Returns cluster tightly around the mean, indicating lower risk.
- In both full and post-1957 data, the **2- and 5-year** horizons have SD > mean (higher short-term risk), while the **10-year** horizon has SD < mean (long-term investing reduces relative volatility).

5. Summary Statistics by Index

```
stats_filt_idx <- all_returns_filt %>%
  group_by(Index, Horizon) %>%
  summarize(
    Mean_Return   = mean(Return_pct),
    Median_Return = median(Return_pct),
    SD_Return     = sd(Return_pct),
    Pct_Positive  = mean(Return_pct > 0) * 100,
    Pct_Negative  = mean(Return_pct < 0) * 100
  )
```

```
## `summarise()` has grouped output by 'Index'. You can override using the
## `.groups` argument.
```

```
knitr::kable(stats_filt_idx, digits = 2,
             caption = "Summary Statistics by Index")
```

Table 2: Summary Statistics by Index

Index	Horizon	Mean_Return	Median_Return	SD_Return	Pct_Positive	Pct_Negative
Dow Jones	2 Years	16.25	15.51	22.17	78.12	21.87
Dow Jones	5 Years	44.99	39.87	46.59	81.82	18.17
Dow Jones	10 Years	110.12	86.79	94.71	88.93	11.07
NASDAQ	2 Years	26.10	26.71	34.58	80.50	19.49
NASDAQ	5 Years	80.13	79.21	74.29	86.80	13.20
NASDAQ	10 Years	209.91	190.53	155.87	95.06	4.94
S&P 500	2 Years	17.70	17.01	22.79	82.15	17.83
S&P 500	5 Years	48.70	48.88	46.87	82.03	17.97
S&P 500	10 Years	116.09	105.53	90.00	92.15	7.85

5.1 Interpreting Post-1957 Summary by Index

- **Dow Jones**: 2-year SD (22.2) > mean (16.2) indicates higher short-term volatility; at 5 years SD > mean (46.6 vs 45) shows moderate risk; at 10 years SD (94.7) < mean (110.1) reflects lower relative

risk and more consistent long-term gains.

- **NASDAQ**: 2-year SD (34.6) > mean (26.0) signals volatile 2-year periods; 5-year SD (74.3) < mean (80.0) suggests improved stability; 10-year SD (155.9) < mean (210) demonstrates strong, reliable long-term performance.

- **S&P 500**: 2-year SD (22.8) > mean (17.7) shows short-term swings; 5-year SD (46.9) slightly < mean (48.7) indicates moderate volatility; 10-year SD (90) < mean (116.1) confirms reduced risk and dependable returns over a decade.

6. Visualizing Distributions of Returns: Full vs Filtered

```
# Compute stats for full data
stats_full <- all_returns %>%
  group_by(Horizon) %>%
  summarize(
    Mean_Return   = mean(Return_pct),
    Median_Return = median(Return_pct),
    Pct_Positive  = mean(Return_pct > 0) * 100
  )

# Plot full data histograms
p1 <- ggplot(all_returns, aes(x = Return_pct)) +
  geom_histogram(bins = 50, fill = 'grey70', color = 'black') +
  geom_vline(data = stats_full, aes(xintercept = Mean_Return),
             color = 'red', linetype = 'dashed', linewidth = 0.7) +
  geom_text(data = stats_full,
            aes(x = Inf, y = Inf,
                label = sprintf("Mean=%.1f%%
Median=%.1f%%
%.0f%%+",
                                Mean_Return, Median_Return, Pct_Positive)),
            hjust = 1.1, vjust = 1.1, size = 7) +
  facet_wrap(~ Horizon, scales = 'free_x') +
  labs(title = 'Full Data: Distribution of Returns') +
  theme_minimal()

# Compute stats for filtered data
stats_filt <- all_returns_filt %>%
  group_by(Horizon) %>%
  summarize(
    Mean_Return   = mean(Return_pct),
    Median_Return = median(Return_pct),
    Pct_Positive  = mean(Return_pct > 0) * 100
  )

# Plot filtered data histograms
p2 <- ggplot(all_returns_filt, aes(x = Return_pct)) +
  geom_histogram(bins = 50, fill = 'steelblue', color = 'black') +
  geom_vline(data = stats_filt, aes(xintercept = Mean_Return),
             color = 'red', linetype = 'dashed', linewidth = 0.7) +
```

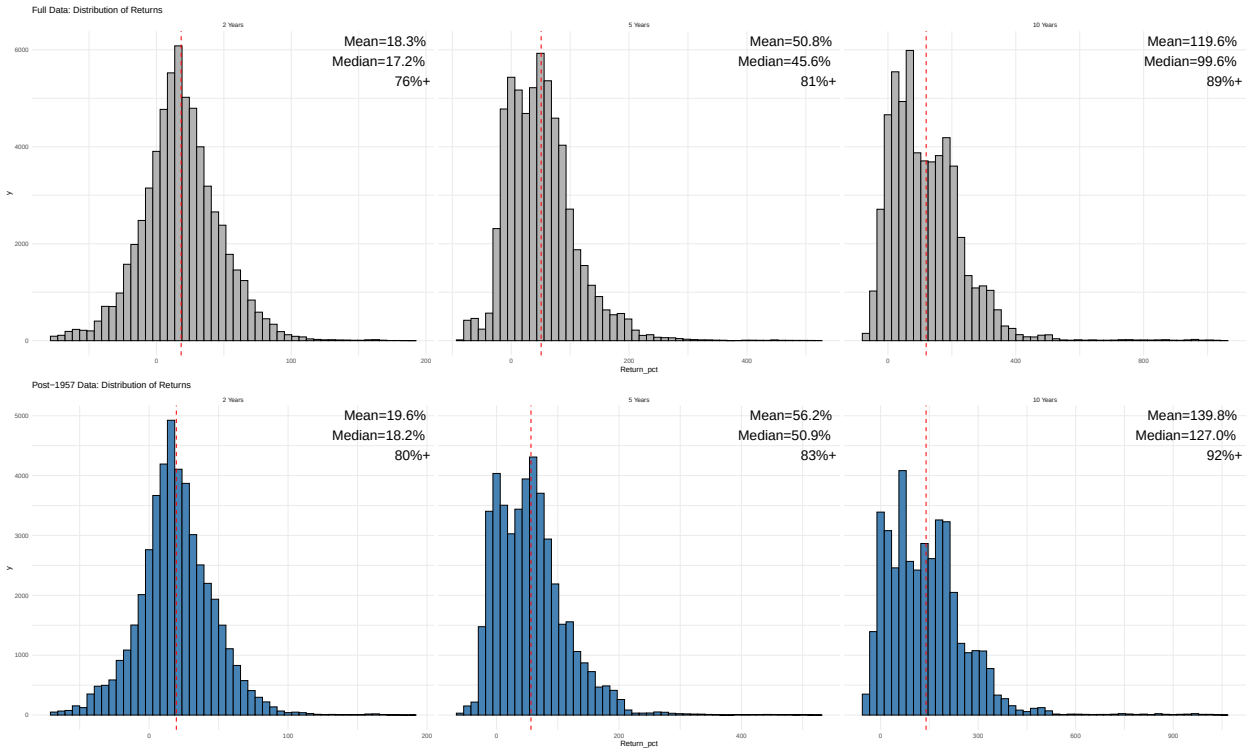
```
  geom_text(data = stats_filt,
            aes(x = Inf, y = Inf,
                label = sprintf("Mean=%.1f%%
Median=%.1f%%
%.0f%%+",
                                Mean_Return, Median_Return, Pct_Positive)),
            hjust = 1.1, vjust = 1.1, size = 7) +
  facet_wrap(~ Horizon, scales = 'free_x') +
  labs(title = 'Post-1957 Data: Distribution of Returns') +
  theme_minimal()

# Print side by side
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

```
grid.arrange(p1, p2, ncol = 1)
```



6.1 Description of Distributions

Full Data

- **2 Years:** Wide distribution with ~24% negative returns; mean ~18%, median ~17%.

- **5 Years:** Bulk shifts right with ~19% negative; mean ~51%, median ~45.5%.
- **10 Years:** Tight cluster with ~11% negative; mean ~119.5%, median ~99.5%.

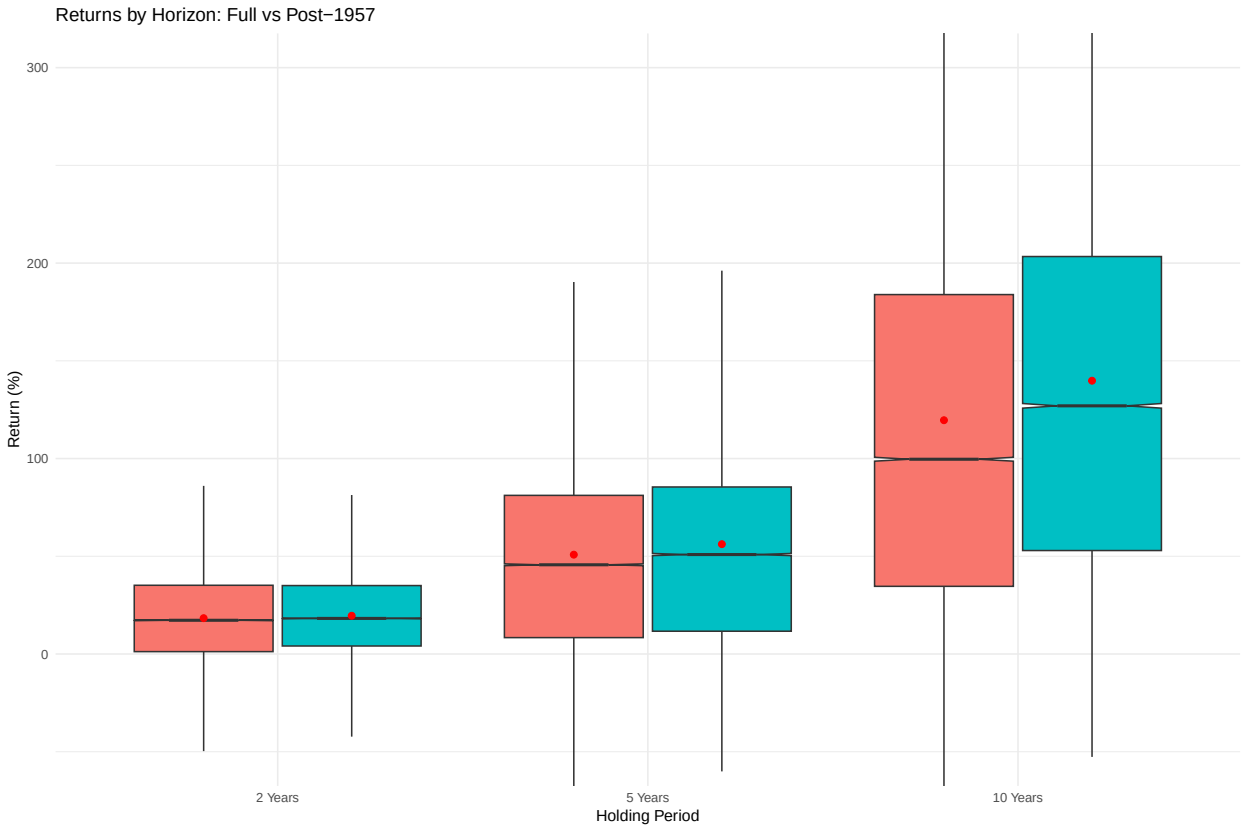
Post-1957 Data

- **2 Years:** Right-shifted, ~20% negative; mean ~19.5%, median ~18%.
- **5 Years:** ~17% negative; mean ~56%, median ~51%.
- **10 Years:** ~8% negative; mean ~140%, median ~127%.

7. Visualizing Relationship: Holding Period vs Return

```
# Boxplot:
#   • notch = approximate 95% CI around the median (±1.58·IQR/sqrt(n))
#   • red dot = mean
#   • y-axis limited to -50%-300% to focus on main distribution
all_compare <- bind_rows(
  all_returns %>% mutate(Scenario = 'Full'),
  all_returns_filt %>% mutate(Scenario = 'Post-1957')
)

ggplot(all_compare, aes(x = Horizon, y = Return_pct, fill = Scenario)) +
  geom_boxplot(notch = TRUE, outlier.shape = NA, position = position_dodge(width = 0.8)) +
  stat_summary(
    fun = mean, geom = "point",
    shape = 20, size = 3, color = "red",
    position = position_dodge(width = 0.8)
  ) +
  coord_cartesian(ylim = c(-50, 300)) +
  labs(
    title = 'Returns by Horizon: Full vs Post-1957',
    x = 'Holding Period', y = 'Return (%)'
  ) +
  theme_minimal() +
  theme(legend.position = 'none')
```



7.1 Explanation of Boxplots

- Notches show 95% confidence intervals around medians—non-overlapping notches imply significantly different medians.
- Red dots mark mean returns.
- By capping the y-axis to -50%–300%, we focus on the core distribution and reduce the visual impact of extreme outliers.

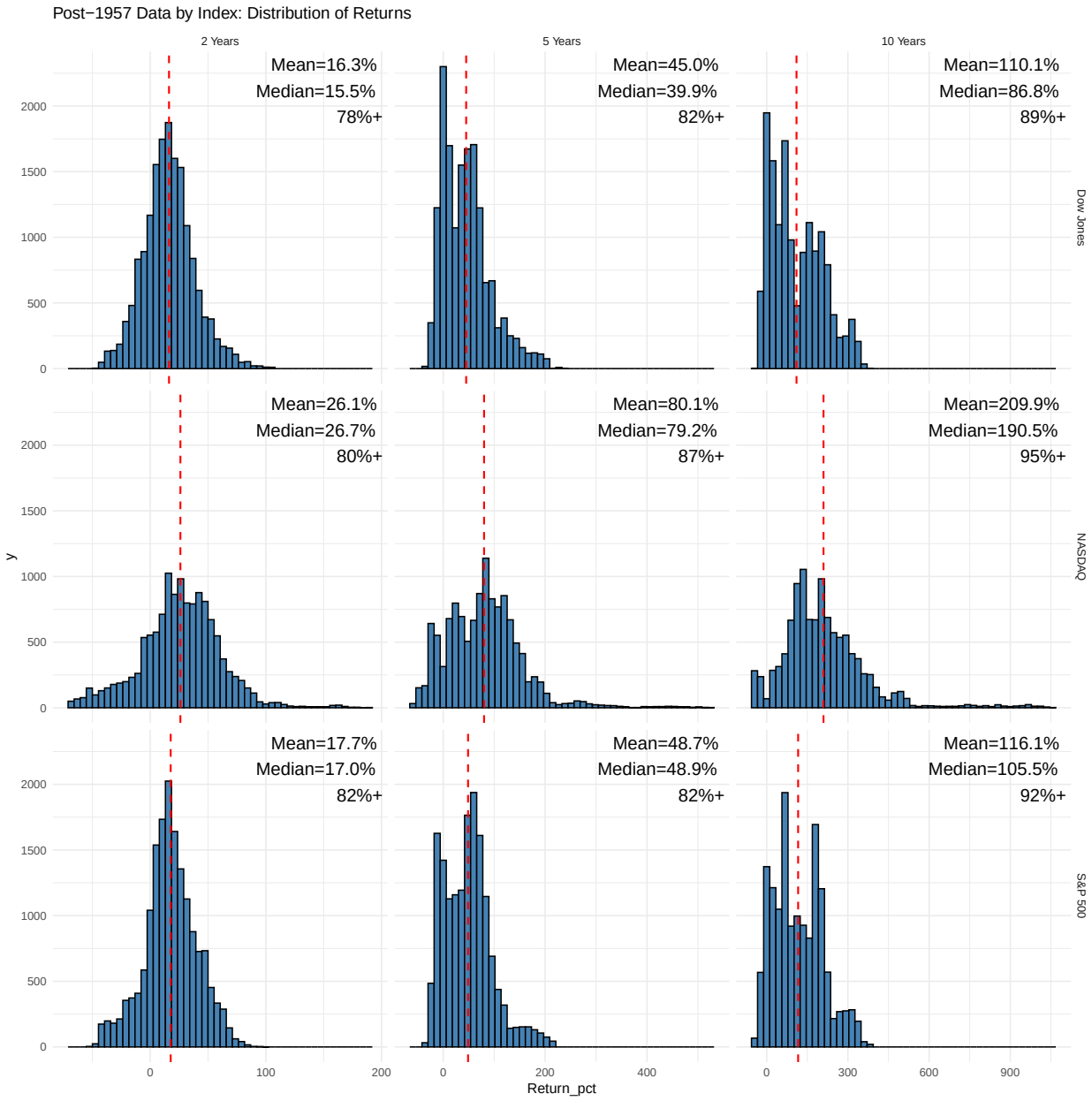
Notice how the post-1957 data consistently shows higher medians and means and tighter interquartile ranges, highlighting reduced relative volatility in the modern market.

8. Histograms by Index

```
# Compute stats by index and horizon
stats_filt_idx <- all_returns_filt %>%
  group_by(Index, Horizon) %>%
  summarize(
    Mean_Return   = mean(Return_pct),
    Median_Return = median(Return_pct),
    Pct_Positive  = mean(Return_pct > 0) * 100
  )
```

```
## `summarise()` has grouped output by 'Index'. You can override using the
## `.groups` argument.
```

```r
# Post-1957 data by index with corner annotations
ggplot(all_returns_filt, aes(x = Return_pct)) +
  geom_histogram(bins = 50, fill = 'steelblue', color = 'black') +
  geom_vline(data = stats_filt_idx, aes(xintercept = Mean_Return),
             color = 'red', linetype = 'dashed', linewidth = 0.7) +
  geom_text(data = stats_filt_idx,
            aes(
              x = Inf, y = Inf,
              label = sprintf("Mean=%.1f%%
Median=%.1f%%
%.0f%%+",
                              Mean_Return, Median_Return, Pct_Positive)
            ),
            hjust = 1.1, vjust = 1.1, size = 5) +
  facet_grid(Index ~ Horizon, scales = 'free_x') +
  labs(title = 'Post-1957 Data by Index: Distribution of Returns') +
  theme_minimal()
```

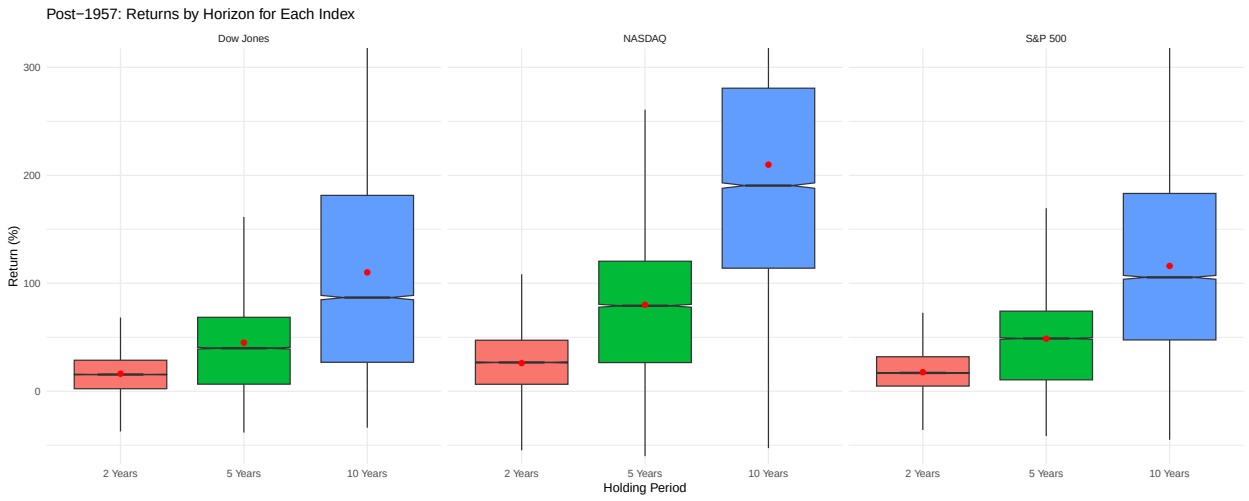


8.1. Interpreting Histograms by Index

- Across indices, the 2-year histograms show wider distributions and greater negative-return percentages, indicating shorter-term volatility.
- The 5-year panels shift right with fewer negatives, illustrating improved performance consistency mid-term.
- The 10-year panels cluster tightly with right-skewed tails, demonstrating strong long-term gains and minimal downside across all indices.
- NASDAQ exhibits the highest means and percentages positive in each horizon, reflecting its growth-oriented nature.

9. Boxplots by Index

```
ggplot(all_returns_filt, aes(x = Horizon, y = Return_pct, fill = Horizon)) +
  geom_boxplot(notch = TRUE, outlier.shape = NA) +
  stat_summary(fun = mean, geom = "point",
               shape = 20, size = 3, color = "red") +
  facet_wrap(~ Index, nrow = 1) +
  coord_cartesian(ylim = c(-50, 300)) +
  labs(
    title = 'Post-1957: Returns by Horizon for Each Index',
    x = 'Holding Period',
    y = 'Return (%)'
  ) +
  theme_minimal() +
  theme(legend.position = 'none')
```



9.1. Interpreting Boxplots by Index

- The notched boxplots confirm median differences across horizons within each index: longer horizons have higher medians.
- Red mean points lie above medians for longer horizons, especially pronounced in NASDAQ's 10-year box, indicating skew.
- Interquartile ranges shrink with holding period, showing reduced volatility over longer terms across all indices.
- NASDAQ's boxes consistently sit higher than Dow and S&P, reaffirming its superior return profile across horizons.

10. Statistical Tests for Horizon and Index

```
# ANOVA: difference in mean returns across holding periods (Horizon)
anova_horizon_full <- aov(Return_pct ~ Horizon, data = all_returns)
anova_horizon_filt <- aov(Return_pct ~ Horizon, data = all_returns_filt)
```

```r
# ANOVA: difference in mean returns across indices (Index)
anova_index_full <- aov(Return_pct ~ Index, data = all_returns)
anova_index_filt <- aov(Return_pct ~ Index, data = all_returns_filt)

# Summaries of tests
summary(anova_horizon_full)
```

```
##                 Df    Sum Sq   Mean Sq F value Pr(>F)
## Horizon          2 315186256 157593128   28435 <2e-16 ***
## Residuals   179871 996880942      5542
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(anova_horizon_filt)
```

```
##                 Df    Sum Sq   Mean Sq F value Pr(>F)
## Horizon          2 323860884 161930442   27282 <2e-16 ***
## Residuals   131214 778800316      5935
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(anova_index_full)
```

```
##                 Df    Sum Sq  Mean Sq F value Pr(>F)
## Index            2 7.433e+07 37167193    5401 <2e-16 ***
## Residuals   179871 1.238e+09     6881
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(anova_index_filt)
```

```
##                 Df    Sum Sq  Mean Sq F value Pr(>F)
## Index            2 4.996e+07 24980764    3114 <2e-16 ***
## Residuals   131214 1.053e+09     8023
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

These ANOVA tests evaluate whether mean returns differ significantly by holding period (Horizon) and by market index (Index) for both the full and post-SPX inception datasets.

10.1. Explanation of Statistical Tests

The ANOVA results show highly significant differences in mean returns:

- **Horizon (Full Dataset):** $F(2, 179871) = 28435$, $p < 2e\text{-}16$ indicates return means differ across 2-, 5-, and 10-year holding periods for the full data.
- **Horizon (Post-SPX Dataset):** $F(2, 131214) = 27282$, $p < 2e\text{-}16$ likewise confirms significant mean differences post-1957.

- **Index (Full Dataset):** F(2, 179871) = 5401, p < 2e-16 shows significant differences in return means among Dow, S&P 500, and NASDAQ for the full data.
- **Index (Post-SPX Dataset):** F(2, 131214) = 3114, p < 2e-16 confirms the same for the post-1957 subset.

All p-values are below conventional significance levels (***), indicating strong evidence that both holding period and market index affect average returns.

11. Annualized Return Analysis

To compare these multi-year returns to typical bank or CD rates, we convert each holding-period return into an **annualized geometric return**:

12. Annualized Summary Statistics by Horizon

```
ann_horizons <- all_ann %>%
  group_by(Horizon) %>%
  summarize(
    Mean_Ann      = mean(Annualized)*100,
    Median_Ann    = median(Annualized)*100,
    SD_Ann        = sd(Annualized)*100,
    Pct_Positive  = mean(Annualized > 0) * 100,
    Pct_Negative  = mean(Annualized < 0) * 100
  )

knitr::kable(
  ann_horizons,
  digits = 2,
  caption = "Table 3: Annualized Return Statistics by Horizon (Post-1957)"
)
```

Table 3: Table 3: Annualized Return Statistics by Horizon (Post-1957)

Horizon	Mean_Ann	Median_Ann	SD_Ann	Pct_Positive	Pct_Negative
2 Years	8.63	8.72	12.47	80.23	19.75
5 Years	8.23	8.57	7.68	83.29	16.70
10 Years	7.97	8.54	5.37	91.79	8.21

12.1: Interpreting Annualized Returns by Horizon

- **2 Years:** Mean = 8.63%, Median = 8.72%, SD = 12.47%; 80.23% of periods are positive, 19.75% negative, indicating shorter horizons retain some downside risk.
- **5 Years:** Mean = 8.23%, Median = 8.57%, SD = 7.68%; 83.29% positive, 16.70% negative, showing improved consistency relative to 2-year returns.
- **10 Years:** Mean = 7.97%, Median = 8.54%, SD = 5.37%; 91.79% positive, 8.21% negative, demonstrating highest reliability and lowest volatility over a decade.

13. Annualized Summary Statistics by Index

```
ann_index <- all_ann %>%
  group_by(Index, Horizon) %>%
  summarize(
    Mean_Ann      = mean(Annualized)*100,
    Median_Ann    = median(Annualized)*100,
    SD_Ann        = sd(Annualized)*100,
    Pct_Positive  = mean(Annualized > 0) * 100,
    Pct_Negative  = mean(Annualized < 0) * 100
  )
```

```
## `summarise()` has grouped output by 'Index'. You can override using the
## `.groups` argument.
```

```
knitr::kable(
  ann_index,
  digits = 2,
  caption = "Table 4: Annualized Return Statistics by Index (Post-1957)"
)
```

Table 4: Table 4: Annualized Return Statistics by Index (Post-1957)

Index	Horizon	Mean_Ann	Median_Ann	SD_Ann	Pct_Positive	Pct_Negative
Dow Jones	2 Years	7.33	7.48	10.30	78.12	21.87
Dow Jones	5 Years	6.88	6.94	6.61	81.82	18.17
Dow Jones	10 Years	6.69	6.45	4.97	88.93	11.07
NASDAQ	2 Years	11.14	12.56	16.08	80.50	19.49
NASDAQ	5 Years	10.99	12.38	9.26	86.80	13.20
NASDAQ	10 Years	10.76	11.25	5.63	95.06	4.94
S&P 500	2 Years	7.96	8.17	10.77	82.15	17.83
S&P 500	5 Years	7.42	8.29	6.71	82.03	17.97
S&P 500	10 Years	7.11	7.47	4.74	92.15	7.85

13.1. Interpreting Annualized Returns by Index

- **Dow Jones:** 2-year (Mean=7.33%, SD=10.30%, 78.12% positive), 5-year (Mean=6.88%, SD=6.61%, 81.82% positive), 10-year (Mean=6.69%, SD=4.97%, 88.93% positive) – consistent improvement and reduced volatility with longer horizons.
- **NASDAQ:** 2-year (Mean=11.14%, SD=16.08%, 80.50% positive), 5-year (Mean=10.99%, SD=9.26%, 86.80% positive), 10-year (Mean=10.76%, SD=5.63%, 95.06% positive) – highest returns and reliability over all horizons.
- **S&P 500:** 2-year (Mean=7.96%, SD=10.77%, 82.15% positive), 5-year (Mean=7.42%, SD=6.71%, 82.03% positive), 10-year (Mean=7.11%, SD=4.74%, 92.15% positive) – moderate returns with strong long-term consistency.

14. Takeaway

The analysis demonstrates that **holding period** and **market index** both play significant roles in U.S. equity returns:

- **Short-term horizons (2–5 years)** carry elevated volatility and a meaningful chance of negative outcomes (15–24%); median returns hover near the average, but risk remains high relative to mean performance.
- **Long-term horizons (10 years)** deliver substantially higher cumulative returns (110–210% raw; 7.97–10.76% annualized) with far lower relative volatility (SD < mean) and negative-return probabilities under 12% (around 8% annualized), affirming the benefits of patience.
- **Post-1957 era** consistently boosts mean and median returns across all horizons and indices, reflecting a more stable, growth-oriented market after S&P 500 inception.
- **Index comparisons** reveal NASDAQ outperforms in both absolute and annualized returns, especially over longer periods, while Dow and S&P 500 show similar patterns of risk reduction but at lower average returns.

Context vs. Traditional Savings Vehicles: Compared to typical cash and fixed-income options, these equity returns are substantial: - **Certificates of Deposit (CDs):** Often yield under 1%–2% annualized.
- **High-Yield Savings Accounts:** Currently offer around 3%–4% APY.
- **Money Market Funds:** Typically yield around 4% APY.

By contrast, even the lowest annualized equity returns (approximately 7.97% over 10 years post-1957) significantly outpace these alternatives, demonstrating the historical advantage of a long-term equity strategy for young investors seeking growth beyond traditional safe-haven rates.

Implications for Younger Investors: For those balancing risk and return—perhaps new savers choosing between savings accounts, CDs, or equities—the data suggest that a diversified, long-horizon equity allocation can offer materially higher returns over time, while still mitigating downside through extended holding periods.