Best Schools for Your Money

Patrick Qi

Table of contents

1 Data Preparation

1.1 Read the Raw Data

The analysis uses two sources:

- `school_data.RData` contains two data frames:
  - `four_year_demog` (1,194 four-year colleges, demographic variables) and
  - `four_year_finan` (1,180 four-year colleges, financial variables).
- `state_info.csv` provides economic and social indicators for all 50 U.S. states.

```
load("school_data.RData")
state_info <- read_csv("state_info.csv", show_col_types = FALSE)
```

1.2 Combine the Data Frames

Three data frames are merged in two steps:

- First, the two school data frames are joined on `UNITID` (unique institutional ID). Because `four_year_finan` has 1,180 records versus 1,194 in `four_year_demog`, an **inner join** retains only the 1,180 schools with complete financial data.
- Second, the combined school data is **left-joined** with `state_info` on `STABBR`, appending state-level context while preserving all schools.

```
school_merged <- merge(four_year_demog, four_year_finan, by = "UNITID")
full_data     <- merge(school_merged, state_info, by = "STABBR", all.x =
↪  TRUE)

cat("Merged dataset:", nrow(full_data), "rows x", ncol(full_data),
↪  "columns\n")
```

```
Merged dataset: 1180 rows x 83 columns
```

1.3 Explore the Raw Structure

Having already inspected the full merged dataset locally, a representative subset of seven columns is shown here to keep the report concise. These columns were selected because they illustrate the storage-type issues present throughout the data; the complete list of corrected columns is reported at the end of the next section.

```r
inspect_cols <- c("INSTNM", "STABBR", "COSTT4_A", "MD_EARN_WNE_1YR",
                  "MD_EARN_WNE_5YR", "DEBT_MDN", "EARN_THR_STATE")
str(full_data[, inspect_cols])
```

```
'data.frame':   1180 obs. of  7 variables:
 $ INSTNM         : chr  "University of Alaska Anchorage" "University of Alaska Fairbanks" "
 $ STABBR         : chr  "AK" "AK" "AL" "AL" ...
 $ COSTT4_A       : chr  "$24,096" "$22,407" "$33,382" "$23,586" ...
 $ MD_EARN_WNE_1YR: chr  "$51,493" "$48,072" "$47,722" "$29,446" ...
 $ MD_EARN_WNE_5YR: chr  "$63,190" "$58,051" "$62,983" "$43,913" ...
 $ DEBT_MDN       : chr  "11231" "11283" "17986" "17500" ...
 $ EARN_THR_STATE : chr  "$40,140" "$40,140" "$30,168" "$30,168" ...
```

The `str()` output confirms the problem:

- Financial variables like `COSTT4_A`, `MD_EARN_WNE_1YR`, and `EARN_THR_STATE` are stored as `chr` (character) rather than `num`, with dollar signs and comma separators embedded in the values.
- `DEBT_MDN` is similarly stored as character even without a dollar sign. These must all be converted to numeric before any arithmetic can proceed.

1.4 Correct the Errors

A custom function, `convert_numeric_like()`, handles the conversion uniformly across all columns in a single pass. The function strips `$`, `,`, and `%` characters and converts empty strings to `NA`, then attempts numeric parsing.

A column is converted only when **at least 50% of its non-missing values parse successfully as numbers**, this threshold preserves genuinely character columns (institution names, city names, state abbreviations, which parse at essentially 0%) while correctly handling financial columns that contain a minority of `"PrivacySuppressed"` entries (which become `NA` after stripping, but the remaining values still parse at well above 50%).

```r
convert_numeric_like <- function(col) {
  if (!is.character(col)) return(col)
  cleaned <- gsub("[$,%]", "", col)
  cleaned[cleaned == ""] <- NA
  non_na <- cleaned[!is.na(cleaned)]
  if (length(non_na) == 0) return(col)
  converted <- suppressWarnings(as.numeric(non_na))
  # Convert if  50% of non-NA values parse as numbers.
  # True text columns (school names, state codes) are ~0% numeric;
  # financial columns with "PrivacySuppressed" entries are still >80%
   ↪  numeric.
  if (mean(!is.na(converted)) < 0.5) return(col)
  suppressWarnings(as.numeric(cleaned))
}

# Record column classes before conversion
classes_before <- sapply(full_data, class)

# Apply to all columns in-place
full_data[] <- lapply(full_data, convert_numeric_like)

# Report which columns were converted
classes_after  <- sapply(full_data, class)
converted_cols <- names(classes_before)[classes_before != classes_after]
cat("Columns converted from character to numeric:\n")
```

```
Columns converted from character to numeric:
```

```r
str(full_data[, converted_cols])
```

```
'data.frame':   1180 obs. of  34 variables:
 $ MD_EARN_WNE_1YR                : num  51493 48072 47722 29446 42602 ...
 $ MD_EARN_WNE_5YR                : num  63190 58051 62983 43913 51119 ...
 $ PCT25_EARN_WNE_P10             : num  28505 25153 39848 18132 28035 ...
 $ PCT75_EARN_WNE_P10             : num  80117 74202 87405 51026 64633 ...
 $ NPT41_PUB                      : num  11582 7992 19169 19711 12472 ...
 $ NPT42_PUB                      : num  13401 6065 19884 21682 12622 ...
 $ NPT43_PUB                      : num  16662 11323 22258 21508 15343 ...
 $ NPT44_PUB                      : num  18826 15526 25658 21608 17480 ...
 $ NPT45_PUB                      : num  20975 17512 26729 19031 18183 ...
 $ NPT41_PRIV                     : num  NA NA NA NA NA ...
```

```
 $ NPT42_PRIV                         : num  NA NA NA NA NA ...
 $ NPT43_PRIV                         : num  NA NA NA NA NA ...
 $ NPT44_PRIV                         : num  NA NA NA NA NA ...
 $ NPT45_PRIV                         : num  NA NA NA NA NA ...
 $ NPT4_PUB                           : num  15301 10892 22420 20435 14279 ...
 $ NPT4_PRIV                          : num  NA NA NA NA NA ...
 $ COSTT4_A                           : num  24096 22407 33382 23586 24257 ...
 $ DEBT_MDN                           : num  11231 11283 17986 17500 14000 ...
 $ TUITIONFEE_IN                      : num  7738 8736 12180 11248 12894 ...
 $ TUITIONFEE_OUT                     : num  21322 22320 34172 19576 23334 ...
 $ TUITFTE                            : num  7357 9659 14144 9820 6358 ...
 $ INEXPFTE                           : num  16384 19769 9728 10723 6997 ...
 $ EARN_THR_STATE                     : num  40140 40140 30168 30168 30168 ...
 $ Total_Employment_Rate_Pct          : num  61.6 61.6 55.5 55.5 55.5 55.5 55.5 55.5 55.5 5
 $ Annual_Change_Pct_Employed         : num  2 2 0.8 0.8 0.8 0.8 0.8 0.8 0.8 0.8 ...
 $ GDP_Pct_Change                     : num  3.6 3.6 4.1 4.1 4.1 4.1 4.1 4.1 4.1 4.1 ...
 $ Income2023                         : num  86631 86631 62212 62212 62212 ...
 $ Income2019                         : num  75463 75463 51734 51734 51734 ...
 $ Income2015                         : num  73355 73355 44765 44765 44765 ...
 $ ResearchExpenditures_Federal_Govt_pct: num  25.7 25.7 25.3 25.3 25.3 25.3 25.3 25.3 25.3 2
 $ ResearchExpenditures_GDP_pct       : num  0.67 0.67 2.54 2.54 2.54 2.54 2.54 2.54 2.54 2
 $ Poverty_Pct                        : num  10.4 10.4 15.6 15.6 15.6 15.6 15.6 15.6 15.6 1
 $ Pct_Forested                       : num  35.2 35.2 70.6 70.6 70.6 ...
 $ Pct_Gun_At_Home                    : num  57 57 53 53 53 53 53 53 53 53 ...
```

1.5 Review the Cleaned Result with Summary Statistics (Tabled)

The cleaning step converted 34 columns from character to numeric. To verify the result and characterize the data before sampling, the table below focuses on the eight variables that drive the financial analysis — six outcome and cost measures plus admission rate and enrollment as institutional context.

```r
key_vars <- c("COSTT4_A", "MD_EARN_WNE_1YR", "MD_EARN_WNE_5YR",
              "EARN_THR_STATE", "PCT25_EARN_WNE_P10", "PCT75_EARN_WNE_P10",
              "ADM_RATE", "UGDS")

var_labels <- c(
  "Annual Cost of Attendance",
  "Median Earnings - 1yr Post-Grad",
  "Median Earnings - 5yr Post-Grad",
  "HS Graduate Earnings (State)",
  "25th Pct Earnings - 10yr",
```

```
  "75th Pct Earnings - 10yr",
  "Admission Rate",
  "Undergraduate Enrollment"
)

eda_summary <- function(df, vars, labels) {
  data.frame(
    Variable = labels,
    N        = sapply(vars, function(v) sum(!is.na(df[[v]]))),
    Missing  = sapply(vars, function(v) sum( is.na(df[[v]]))),
    Mean     = sapply(vars, function(v) round(mean(  df[[v]], na.rm = TRUE),
     ↪  1)),
    SD       = sapply(vars, function(v) round(sd(    df[[v]], na.rm = TRUE),
     ↪  1)),
    Median   = sapply(vars, function(v) round(median(df[[v]], na.rm = TRUE),
     ↪  1)),
    Min      = sapply(vars, function(v) round(min(   df[[v]], na.rm = TRUE),
     ↪  1)),
    Max      = sapply(vars, function(v) round(max(   df[[v]], na.rm = TRUE),
     ↪  1))
  )
}

kable(
  eda_summary(full_data, key_vars, var_labels),
  row.names   = FALSE,
  caption     = "Descriptive Statistics for Key Analysis Variables",
  col.names   = c("Variable", "N", "Missing", "Mean", "SD", "Median", "Min",
   ↪  "Max"),
  align       = c("l", rep("r", 7)),
  booktabs    = TRUE,
  format.args = list(big.mark = ",")
)
```

Table 1: Descriptive Statistics for Key Analysis Variables

Variable	N	Missing	Mean	SD	Median	Min	Max
Annual Cost of Attendance	1,169	11	44,943.1	20,373.0	42,471.0	11,837	93,512
Median Earnings - 1yr Post-Grad	1,166	14	45,440.8	11,560.8	43,685.0	12,695	109,923

Variable	N	Missing	Mean	SD	Median	Min	Max
Median Earnings - 5yr Post-Grad	1,168	12	60,719.4	13,993.5	58,148.5	27,062	154,095
HS Graduate Earnings (State)	1,180	0	33,167.9	2,004.6	32,990.0	29,193	40,140
25th Pct Earnings - 10yr	1,161	19	37,610.6	10,663.4	36,585.0	10,062	90,680
75th Pct Earnings - 10yr	1,164	16	85,163.0	22,799.0	80,728.0	39,790	243,025
Admission Rate	1,113	67	0.7	0.2	0.8	0	1
Undergraduate Enrollment	1,180	0	6,382.0	10,619.0	2,629.0	23	163,164

1.6 Sample

Using the last four digits of my student ID (**3915**) as the random seed, a reproducible 50% random sample is drawn.

UCLA (UNITID 110662) is extracted from the full dataset independently, regardless of whether it falls into the random sample and is used as a reference benchmark throughout the analysis.

This extra benchmarking of UCLA is a self-directed addition not required by the assignment.

```
set.seed(3915)

n_half      <- floor(nrow(full_data) * 0.5)
sample_idx  <- sample(nrow(full_data), n_half)
sample_data <- full_data[sample_idx, ]

# Extract UCLA from the full dataset independently of the sample
ucla_data <- full_data[full_data$INSTNM == "University of California-Los
 ↪  Angeles", ]

cat("Total cleaned rows:", nrow(full_data), "\n")
```

```
Total cleaned rows: 1180
```

```
cat("50% sample size:   ", nrow(sample_data), "\n")
```

```
50% sample size:    590
```

```r
cat("UCLA extracted from full dataset for reference (independent of
↪  sample).\n")
```

```
UCLA extracted from full dataset for reference (independent of sample).
```

2 Analysis

The three sections below compute a distinct financial metric across the 50% random sample and present the five best- and worst-performing schools.

In addition to the required sample analysis, UCLA's metrics are reported separately after each set of tables as a benchmarking reference.

UCLA is always drawn from the full dataset using identical formulas, so its reference values are consistent regardless of whether it appears in the random sample (with seed 3915, UCLA is not in the sample). Reference rows are marked † to distinguish them from sample results.

Cross-regional note: Comparing schools across states involves inherent limitations, cost of living, local labor-market structure, and state economic policy all vary. The wage premium metric partially addresses this by using a state-specific earnings baseline (`EARN_THR_STATE`), but cross-state comparisons still have real limits.

```
# Compute all analysis metrics for UCLA in one place
ucla_data <- ucla_data %>%
  mutate(
    total_cost_4yr    = 4 * COSTT4_A,
    payback_period    = total_cost_4yr / MD_EARN_WNE_1YR,
    wage_premium_5yr  = MD_EARN_WNE_5YR  - EARN_THR_STATE,
    earn_10yr_mid     = (PCT25_EARN_WNE_P10 + PCT75_EARN_WNE_P10) / 2,
    wage_premium_10yr = earn_10yr_mid - EARN_THR_STATE,
    college_roi       = (5 * wage_premium_5yr - total_cost_4yr) /
                         total_cost_4yr * 100
  )
```

2.1 The Payback Period

Assumptions. The payback period measures how many years of post-graduation earnings are required to fully recoup the cost of a four-year degree.

Cost: `COSTT4_A` — total cost of attendance without scholarships or grants, applied for exactly four years: $4 \times \text{COSTT4\_A}$.

Earnings: `MD_EARN_WNE_1YR` — median earnings of employed graduates one year after completing their degree.

$$\text{Payback Period (years)} = \frac{4 \times \text{COSTT4\_A}}{\text{MD\_EARN\_WNE\_1YR}}$$

Schools missing either variable are excluded.

```r
sample_data <- sample_data %>%
  mutate(
    total_cost_4yr = 4 * COSTT4_A,
    payback_period = total_cost_4yr / MD_EARN_WNE_1YR
  )

payback_df <- sample_data %>%
  filter(!is.na(payback_period) & payback_period > 0) %>%
  arrange(payback_period) %>%
  transmute(
    Institution   = INSTNM,
    State         = STABBR,
    `Cost (USD)`  = round(COSTT4_A, 0),
    `Earn 1yr`    = round(MD_EARN_WNE_1YR, 0),
    `Payback yrs` = round(payback_period, 2)
  )
```

```r
ucla_pb <- data.frame(
  Institution   = paste0(ucla_data$INSTNM, " †"),
  State         = ucla_data$STABBR,
  `Cost (USD)`  = round(ucla_data$COSTT4_A, 0),
  `Earn 1yr`    = round(ucla_data$MD_EARN_WNE_1YR, 0),
  `Payback yrs` = round(ucla_data$payback_period, 2),
  check.names   = FALSE
)
pb_sep <- data.frame(
  Institution = "---", State = "---",
  `Cost (USD)` = NA, `Earn 1yr` = NA, `Payback yrs` = NA,
  check.names = FALSE
)
```

The five schools with the **shortest** payback periods (shortest to longest), with UCLA shown below the divider for reference († full-dataset value, independent of sample):

```r
kable(
  rbind(head(payback_df, 5), pb_sep, ucla_pb),
  caption     = "5 Schools with Shortest Payback Period (Sample) + UCLA
  ↪  Reference",
  row.names   = FALSE,
  booktabs    = TRUE,
  format.args = list(big.mark = ",")
)
```

Table 2: 5 Schools with Shortest Payback Period (Sample) + UCLA Reference

Institution	State	Cost (USD)	Earn 1yr	Payback yrs
Western Governors University	UT	16,728	74,978	0.89
Indiana University-Kokomo	IN	11,837	43,591	1.09
CUNY Bernard M Baruch College	NY	14,170	51,920	1.09
Texas A&M University-Central Texas	TX	14,187	45,636	1.24
Georgia Institute of Technology-Main Campus	GA	28,167	87,556	1.29
—	—			
University of California-Los Angeles †	CA	38,614	41,763	3.70

The five schools with the **longest** payback periods (longest to shortest), with UCLA shown below the divider for reference:

```
kable(
  rbind(tail(payback_df, 5) %>% arrange(desc(`Payback yrs`)), pb_sep,
   ↪ ucla_pb),
  caption     = "5 Schools with Longest Payback Period (Sample) + UCLA
   ↪ Reference",
  row.names   = FALSE,
  booktabs    = TRUE,
  format.args = list(big.mark = ",")
)
```

Table 3: 5 Schools with Longest Payback Period (Sample) + UCLA Reference

Institution	State	Cost (USD)	Earn 1yr	Payback yrs
Bard College	NY	84,553	21,735	15.56
Hampshire College	MA	74,195	20,377	14.56
Rhode Island School of Design	RI	81,727	22,778	14.35
Berklee College of Music	MA	72,700	20,342	14.30
The New School	NY	88,284	26,730	13.21
—	—			
University of California-Los Angeles †	CA	38,614	41,763	3.70

2.2 The Wage Premium

Methodology. A degree's true financial value is the *additional* earnings above what a student would have made without attending college.

I use `EARN_THR_STATE`, median earnings of high-school graduates aged 18–30 in the state where each college is located as the state-specific baseline.

5-year wage premium uses `MD_EARN_WNE_5YR`, median earnings five years after graduation, which better reflects a graduate's settled career earnings:

$$\text{Premium}_{5\text{yr}} = \text{MD\_EARN\_WNE\_5YR} - \text{EARN\_THR\_STATE}$$

10-year wage premium uses the midpoint of the 25th and 75th percentile earnings ten years after entry (approximately age 30, per the variable definition in the codebook):

$$\text{Premium}_{10\text{yr}} = \frac{\text{PCT25\_EARN\_WNE\_P10} + \text{PCT75\_EARN\_WNE\_P10}}{2} - \text{EARN\_THR\_STATE}$$

```
sample_data <- sample_data %>%
  mutate(
    wage_premium_5yr  = MD_EARN_WNE_5YR - EARN_THR_STATE,
    earn_10yr_mid     = (PCT25_EARN_WNE_P10 + PCT75_EARN_WNE_P10) / 2,
    wage_premium_10yr = earn_10yr_mid - EARN_THR_STATE
  )

wp5_df <- sample_data %>%
  filter(!is.na(wage_premium_5yr)) %>%
  arrange(wage_premium_5yr) %>%
  transmute(
    Institution   = INSTNM,
    State         = STABBR,
    `Earn 5yr`    = round(MD_EARN_WNE_5YR, 0),
    `HS Baseline` = round(EARN_THR_STATE, 0),
    `Premium 5yr` = round(wage_premium_5yr, 0)
  )

wp10_df <- sample_data %>%
  filter(!is.na(wage_premium_10yr)) %>%
  arrange(wage_premium_10yr) %>%
  transmute(
    Institution    = INSTNM,
    State          = STABBR,
```

```
    `Earn 10yr mid`= round(earn_10yr_mid, 0),
    `HS Baseline`  = round(EARN_THR_STATE, 0),
    `Premium 10yr` = round(wage_premium_10yr, 0)
  )
```

```
ucla_wp5 <- data.frame(
  Institution   = paste0(ucla_data$INSTNM, " †"),
  State         = ucla_data$STABBR,
  `Earn 5yr`    = round(ucla_data$MD_EARN_WNE_5YR, 0),
  `HS Baseline` = round(ucla_data$EARN_THR_STATE, 0),
  `Premium 5yr` = round(ucla_data$wage_premium_5yr, 0),
  check.names   = FALSE
)
wp5_sep <- data.frame(
  Institution = "---", State = "---",
  `Earn 5yr` = NA, `HS Baseline` = NA, `Premium 5yr` = NA,
  check.names = FALSE
)

ucla_wp10 <- data.frame(
  Institution    = paste0(ucla_data$INSTNM, " †"),
  State          = ucla_data$STABBR,
  `Earn 10yr mid`= round(ucla_data$earn_10yr_mid, 0),
  `HS Baseline`  = round(ucla_data$EARN_THR_STATE, 0),
  `Premium 10yr` = round(ucla_data$wage_premium_10yr, 0),
  check.names    = FALSE
)
wp10_sep <- data.frame(
  Institution = "---", State = "---",
  `Earn 10yr mid` = NA, `HS Baseline` = NA, `Premium 10yr` = NA,
  check.names = FALSE
)
```

5-Year Wage Premium — Largest 5 (highest to lowest), with UCLA reference below divider (†):

```
kable(
  rbind(tail(wp5_df, 5) %>% arrange(desc(`Premium 5yr`)), wp5_sep, ucla_wp5),
  caption     = "5 Schools with Largest 5-Year Wage Premium (Sample) + UCLA
   ↪ Reference",
  row.names   = FALSE, booktabs = TRUE, format.args = list(big.mark = ",")
)
```

Table 4: 5 Schools with Largest 5-Year Wage Premium (Sample) + UCLA Reference

Institution	State	Earn 5yr	HS Baseline	Premium 5yr
Harvey Mudd College	CA	154,095	34,672	119,423
California Institute of Technology	CA	136,775	34,672	102,103
Stanford University	CA	132,900	34,672	98,228
Carnegie Mellon University	PA	121,367	33,196	88,171
Princeton University	NJ	113,450	34,809	78,641
—	—			
University of California-Los Angeles †	CA	77,498	34,672	42,826

5-Year Wage Premium — Smallest 5 (lowest to highest), with UCLA reference:

```
kable(
  rbind(head(wp5_df, 5), wp5_sep, ucla_wp5),
  caption     = "5 Schools with Smallest 5-Year Wage Premium (Sample) + UCLA
  ↪  Reference",
  row.names   = FALSE, booktabs = TRUE, format.args = list(big.mark = ",")
)
```

Table 5: 5 Schools with Smallest 5-Year Wage Premium (Sample) + UCLA Reference

Institution	State	Earn 5yr	HS Baseline	Premium 5yr
Hampshire College	MA	32,244	37,113	-4,869
Berklee College of Music	MA	32,591	37,113	-4,522
Maharishi International University	IA	32,465	34,809	-2,344
Allen University	SC	35,418	31,959	3,459
Prescott College	AZ	38,252	34,059	4,193
—	—			
University of California-Los Angeles †	CA	77,498	34,672	42,826

10-Year Wage Premium — Largest 5 (highest to lowest), with UCLA reference below divider (†):

```
kable(
  rbind(tail(wp10_df, 5) %>% arrange(desc(`Premium 10yr`)), wp10_sep,
  ↪  ucla_wp10),
```

```
  caption     = "5 Schools with Largest 10-Year Wage Premium (Sample) + UCLA
  ↪  Reference",
  row.names   = FALSE, booktabs = TRUE, format.args = list(big.mark = ",")
)
```

Table 6: 5 Schools with Largest 10-Year Wage Premium (Sample) + UCLA Reference

Institution	State	Earn 10yr mid	HS Baseline	Premium 10yr
California Institute of Technology	CA	151,782	34,672	117,110
Stanford University	CA	146,530	34,672	111,858
Harvey Mudd College	CA	143,696	34,672	109,024
Princeton University	NJ	132,904	34,809	98,094
Carnegie Mellon University	PA	127,664	33,196	94,468
—	—			
University of California-Los Angeles †	CA	91,174	34,672	56,502

10-Year Wage Premium — Smallest 5 (lowest to highest), with UCLA reference:

```
kable(
  rbind(head(wp10_df, 5), wp10_sep, ucla_wp10),
  caption     = "5 Schools with Smallest 10-Year Wage Premium (Sample) + UCLA
  ↪  Reference",
  row.names   = FALSE, booktabs = TRUE, format.args = list(big.mark = ",")
)
```

Table 7: 5 Schools with Smallest 10-Year Wage Premium (Sample) + UCLA Reference

Institution	State	Earn 10yr mid	HS Baseline	Premium 10yr
East-West University	IL	29,856	32,990	-3,134
Allen University	SC	30,252	31,959	-1,708
Maharishi International University	IA	33,848	34,809	-961
Berklee College of Music	MA	37,038	37,113	-74
Lane College	TN	31,676	31,130	546
—	—			
University of California-Los Angeles †	CA	91,174	34,672	56,502

2.3 The Cumulative Return on Investment

Methodology. The 5-year cumulative ROI captures the total net financial gain over the first five working years, expressed as a percentage of the degree cost.

5-year cumulative wage premium: $5 \times \text{Premium}_{5\text{yr}}$, representing five years of earnings advantage over a high-school graduate.

Cost of a four-year degree: $4 \times \text{COSTT4\_A}$.

$$\text{CollegeROI} = \frac{5 \times \text{Premium}_{5\text{yr}} - 4 \times \text{COSTT4\_A}}{4 \times \text{COSTT4\_A}} \times 100$$

A positive ROI means the degree more than pays for itself within five working years; a negative ROI means the investment has not yet broken even.

```
sample_data <- sample_data %>%
  mutate(
    college_roi = (5 * wage_premium_5yr - total_cost_4yr) / total_cost_4yr *
     ↪  100
  )

roi_df <- sample_data %>%
  filter(!is.na(college_roi)) %>%
  arrange(college_roi) %>%
  transmute(
    Institution    = INSTNM,
    State          = STABBR,
    `4yr Cost`     = round(total_cost_4yr, 0),
    `Premium 5yr`  = round(wage_premium_5yr, 0),
    `ROI pct`      = round(college_roi, 1)
  )
```

```
ucla_roi <- data.frame(
  Institution    = paste0(ucla_data$INSTNM, " †"),
  State          = ucla_data$STABBR,
  `4yr Cost`     = round(ucla_data$total_cost_4yr, 0),
  `Premium 5yr`  = round(ucla_data$wage_premium_5yr, 0),
  `ROI pct`      = round(ucla_data$college_roi, 1),
  check.names    = FALSE
)
roi_sep <- data.frame(
  Institution = "---", State = "---",
```

```
  `4yr Cost` = NA, `Premium 5yr` = NA, `ROI pct` = NA,
  check.names = FALSE
)
```

Highest 5-Year CollegeROI (highest to lowest), with UCLA reference below divider (†):

```
kable(
  rbind(tail(roi_df, 5) %>% arrange(desc(`ROI pct`)), roi_sep, ucla_roi),
  caption     = "5 Schools with Highest 5-Year CollegeROI (Sample) + UCLA
   ↪  Reference",
  row.names   = FALSE, booktabs = TRUE, format.args = list(big.mark = ",")
)
```

Table 8: 5 Schools with Highest 5-Year CollegeROI (Sample) + UCLA Reference

Institution	State	4yr Cost	Premium 5yr	ROI pct
CUNY Bernard M Baruch College	NY	56,680	49,092	333.1
Western Governors University	UT	66,912	47,237	253.0
Georgia Institute of Technology-Main Campus	GA	112,668	73,995	228.4
CUNY City College	NY	57,612	35,023	204.0
SUNY Maritime College	NY	119,144	70,591	196.2
—	—			
University of California-Los Angeles †	CA	154,456	42,826	38.6

Lowest 5-Year CollegeROI (lowest to highest), with UCLA reference:

```
kable(
  rbind(head(roi_df, 5), roi_sep, ucla_roi),
  caption     = "5 Schools with Lowest 5-Year CollegeROI (Sample) + UCLA
   ↪  Reference",
  row.names   = FALSE, booktabs = TRUE, format.args = list(big.mark = ",")
)
```

Table 9: 5 Schools with Lowest 5-Year CollegeROI (Sample) + UCLA Reference

Institution	State	4yr Cost	Premium 5yr	ROI pct
Maharishi International University	IA	111,568	-2,344	-110.5
Hampshire College	MA	296,780	-4,869	-108.2
Berklee College of Music	MA	290,800	-4,522	-107.8
Bard College	NY	338,212	5,563	-91.8
Prescott College	AZ	194,128	4,193	-89.2
—	—			
University of California-Los Angeles †	CA	154,456	42,826	38.6

2.4 The Context (a Visualization)

A college's ROI does not exist in a vacuum. Even a high-ROI school carries elevated risk if the surrounding job market is weak. I use `SA_UnemploymentRate` (state seasonally-adjusted unemployment rate) as a job-market risk factor: graduates in high-unemployment states face a lower probability of realizing the earnings that drive the ROI calculation. This directly challenges the assumption embedded in the ROI formula: T**hat graduates will find employment.**

An **unemployment-adjusted ROI** penalizes schools in high-unemployment states by up to 40%. The unemployment rate is min-max normalized to $[0, 1]$ across the sample:

$$\text{Adjusted ROI} = \text{CollegeROI} - 0.40 \times \hat{U} \times |\text{CollegeROI}|$$

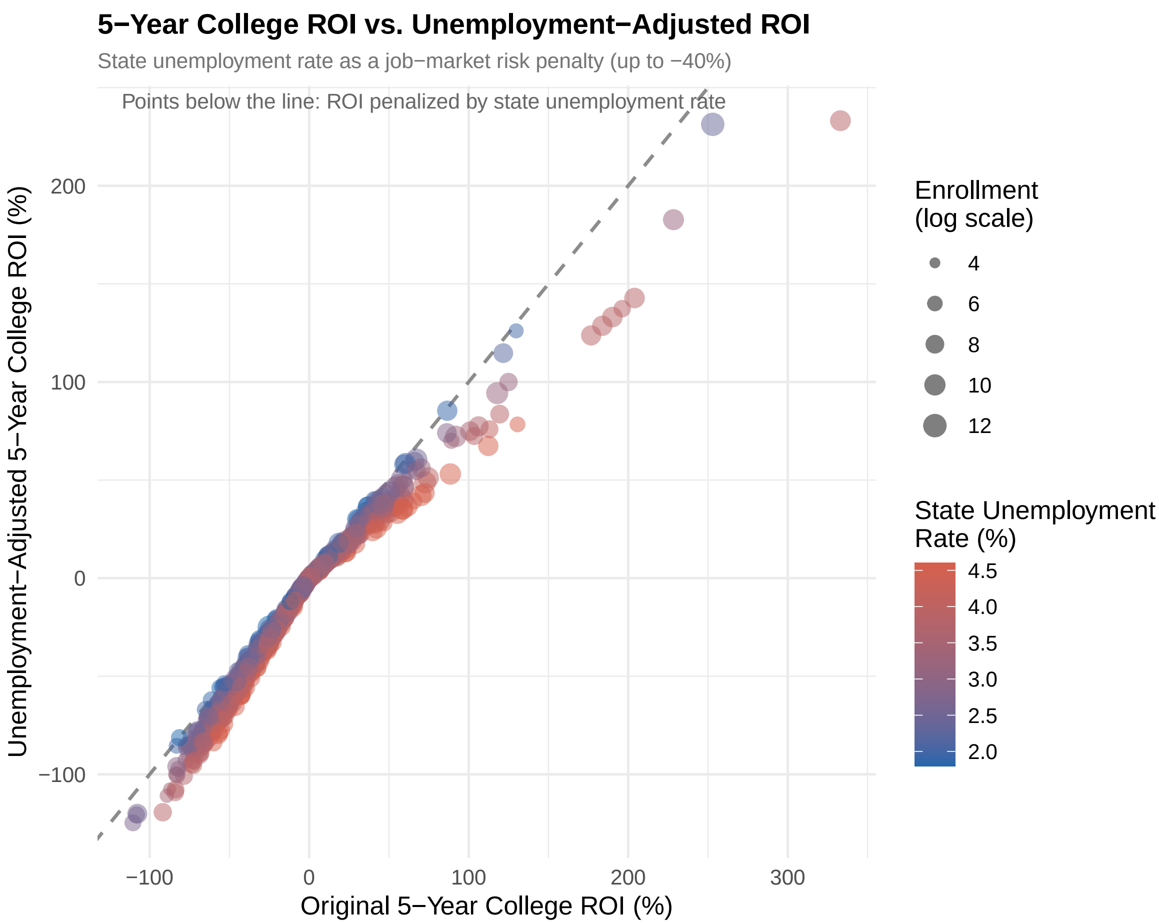
where $\hat{U}$ is the normalized state unemployment rate. The penalty term always moves the adjusted ROI in the unfavorable direction: positive returns are reduced and negative returns are amplified. A naïve multiplicative form $\text{CollegeROI} \times (1 - 0.40\hat{U})$ would produce the opposite effect for negative-ROI schools (high unemployment would make them appear *less* bad), which is logically inconsistent with the intent. Points **below** the 45° reference line represent schools whose nominal ROI overstates the job-market-adjusted return.

Note on excluded observations:

- `state_info` covers all 50 U.S. states but does not include Washington, D.C. Schools located in D.C. therefore have `SA_UnemploymentRate = NA` after the left join and are excluded from this visualization by the `filter(!is.na SA_UnemploymentRate))` call.
- Additionally, the min-max normalization of $\hat{U}$ depends on the sample's unemployment range, so the exact penalty magnitudes will vary across random seeds.

```
viz_data <- sample_data %>%
  filter(!is.na(college_roi) & !is.na(SA_UnemploymentRate)) %>%
  mutate(
    u_norm  = (SA_UnemploymentRate - min(SA_UnemploymentRate, na.rm = TRUE))
     ↪  /
              (max(SA_UnemploymentRate, na.rm = TRUE) -
               min(SA_UnemploymentRate, na.rm = TRUE)),
    adj_roi = college_roi - 0.40 * u_norm * abs(college_roi)
  )

ggplot(viz_data,
       aes(x = college_roi, y = adj_roi,
           color = SA_UnemploymentRate, size = log1p(UGDS))) +
  geom_abline(slope = 1, intercept = 0,
              linetype = "dashed", color = "gray55", linewidth = 0.7) +
  geom_point(alpha = 0.5) +
  scale_color_gradient(
    low  = "#2166AC", high = "#D6604D",
    name = "State Unemployment\nRate (%)"
  ) +
  scale_size_continuous(
    range = c(0.4, 4),
    name  = "Enrollment\n(log scale)"
  ) +
  annotate("text", x = -Inf, y = Inf,
           label = "Points below the line: ROI penalized by state
            ↪  unemployment rate",
           hjust = -0.04, vjust = 1.5, size = 3, color = "gray40") +
  labs(
    title    = "5-Year College ROI vs. Unemployment-Adjusted ROI",
    subtitle = "State unemployment rate as a job-market risk penalty (up to
     ↪  -40%)",
    x        = "Original 5-Year College ROI (%)",
    y        = "Unemployment-Adjusted 5-Year College ROI (%)"
  ) +
  theme_minimal(base_size = 11) +
  theme(
    plot.title      = element_text(face = "bold", size = 12),
    plot.subtitle   = element_text(color = "gray45", size = 9),
    legend.position = "right"
  )
```



5−Year College ROI vs. Unemployment−Adjusted ROI
State unemployment rate as a job−market risk penalty (up to −40%)
Points below the line: ROI penalized by state unemployment rate
Unemployment−Adjusted 5−Year College ROI (%)
200
100
0
−100
−100
0
100
200
300
Original 5−Year College ROI (%)
Enrollment
(log scale)
4
6
8
10
12
State Unemployment
Rate (%)
4.5
4.0
3.5
3.0
2.5
2.0

3 UCLA: An Extended Look

This section extends beyond the core project requirements. The follows is a self-directed exploration of UCLA (University of California, Los Angeles), motivated by UCLA is my alma mater, and this section is a chance to take a closer look at her numbers.

A note on cross-regional comparisons. Section 2 compares schools across all U.S. states. While the wage premium uses a state-specific earnings baseline to partially control for regional differences, factors such as cost of living, industry composition, and state economic policy remain uncontrolled. The comparisons in this section are therefore descriptive — they should be read as "how does UCLA's measured profile compare to this benchmark?" rather than as causal claims. Within-state comparisons (Section 3.1.3) avoid this limitation entirely.

3.1 UCLA at a Glance

3.1.1 Student Composition

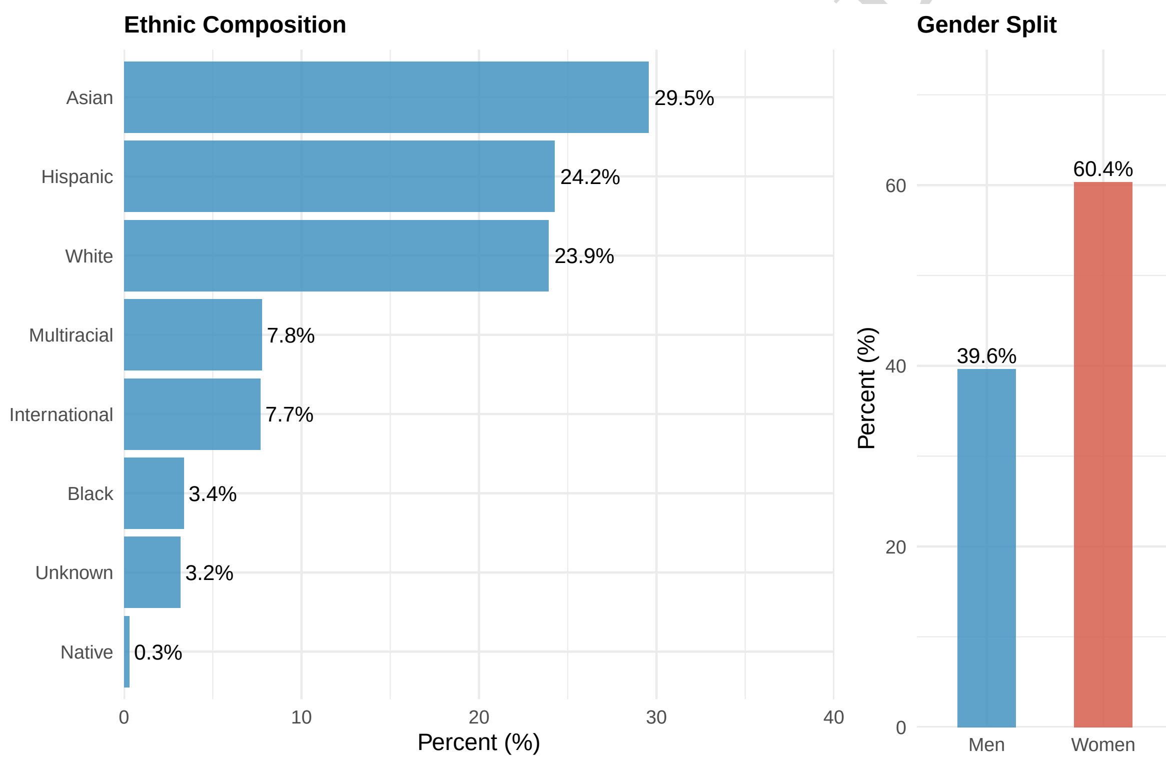
```
eth_df <- data.frame(
  Group = c("Asian", "Hispanic", "White", "Multiracial",
            "International", "Black", "Unknown", "Native"),
  Pct   = c(ucla_data$UGDS_ASIAN, ucla_data$UGDS_HISP,
            ucla_data$UGDS_WHITE, ucla_data$UGDS_2MOR,
            ucla_data$UGDS_NRA,   ucla_data$UGDS_BLACK,
            ucla_data$UGDS_UNKN,
            ucla_data$UGDS_AIAN + ucla_data$UGDS_NHPI) * 100
) %>% arrange(Pct)

p_eth <- ggplot(eth_df, aes(x = reorder(Group, Pct), y = Pct)) +
  geom_col(fill = "#4393C3", alpha = 0.85) +
  geom_text(aes(label = paste0(round(Pct, 1), "%")),
            hjust = -0.1, size = 3) +
  coord_flip() +
  scale_y_continuous(limits = c(0, 40), expand = c(0, 0)) +
  labs(title = "Ethnic Composition", x = NULL, y = "Percent (%)") +
  theme_minimal(base_size = 10) +
  theme(plot.title = element_text(face = "bold", size = 10))

gen_df <- data.frame(
  Gender = c("Women", "Men"),
  Pct    = c(ucla_data$UGDS_WOMEN, ucla_data$UGDS_MEN) * 100
)
```

```
p_gen <- ggplot(gen_df, aes(x = Gender, y = Pct, fill = Gender)) +
  geom_col(alpha = 0.85, width = 0.5) +
  geom_text(aes(label = paste0(round(Pct, 1), "%")),
            vjust = -0.4, size = 3.2) +
  scale_fill_manual(values = c("Women" = "#D6604D", "Men" = "#4393C3")) +
  scale_y_continuous(limits = c(0, 75), expand = c(0, 0)) +
  labs(title = "Gender Split", x = NULL, y = "Percent (%)") +
  theme_minimal(base_size = 10) +
  theme(plot.title = element_text(face = "bold", size = 10),
        legend.position = "none")

grid.arrange(p_eth, p_gen, ncol = 2, widths = c(2.5, 1))
```



3.1.2 Family Economic Background

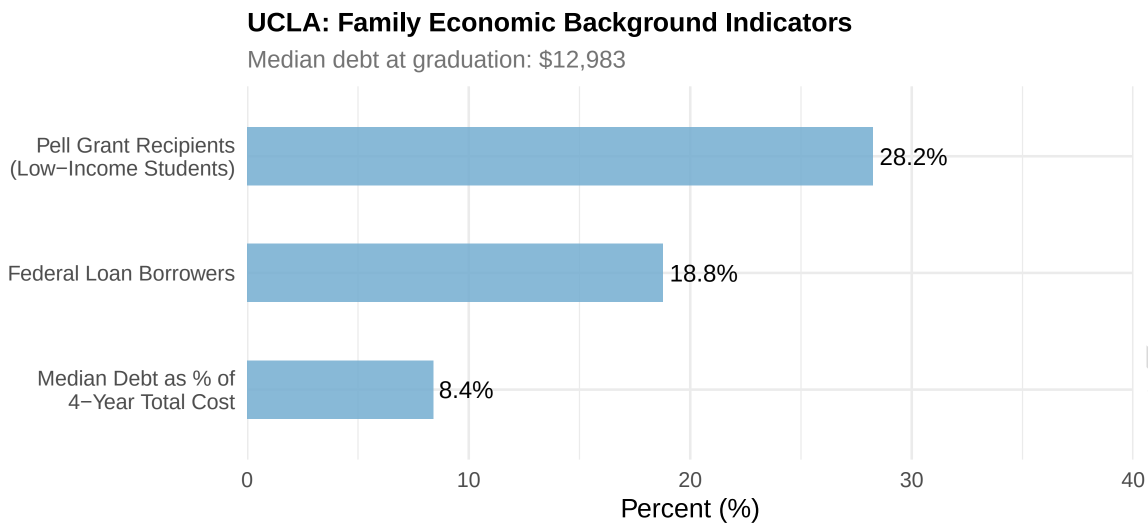
The three indicators below characterize the socioeconomic composition of UCLA's student body.

- Pell Grant recipients are students whose family income typically falls below $60,000 — a direct measure of low-income access.
- The median amount borrowed was only a small fraction of the total cost of attendance — reflecting the impact of grants and institutional aid for most students.
- The debt-as-cost ratio shows that among the 18.8% of students who took out federal loans.

```
debt_pct <- ucla_data$DEBT_MDN / ucla_data$total_cost_4yr * 100

econ_df <- data.frame(
  Indicator = c("Pell Grant Recipients\n(Low-Income Students)",
                "Federal Loan Borrowers",
                "Median Debt as % of\n4-Year Total Cost"),
  Pct = c(ucla_data$PCTPELL  * 100,
          ucla_data$PCTFLOAN * 100,
          debt_pct)
) %>% arrange(Pct)

ggplot(econ_df, aes(x = reorder(Indicator, Pct), y = Pct)) +
  geom_col(fill = "#74ADD1", alpha = 0.85, width = 0.5) +
  geom_text(aes(label = paste0(round(Pct, 1), "%")),
            hjust = -0.1, size = 3.2) +
  coord_flip() +
  scale_y_continuous(limits = c(0, 40), expand = c(0, 0)) +
  labs(
    title    = "UCLA: Family Economic Background Indicators",
    subtitle = paste0("Median debt at graduation: $",
                      format(round(ucla_data$DEBT_MDN), big.mark = ",")),
    x = NULL, y = "Percent (%)"
  ) +
  theme_minimal(base_size = 10) +
  theme(plot.title    = element_text(face = "bold", size = 10),
        plot.subtitle = element_text(color = "gray45", size = 8.5))
```



3.1.3 UCLA vs. California Public Universities

The following chart compares UCLA to the median of all other California public four-year institutions on four key metrics. This within-state comparison avoids the cross-regional limitations discussed above.

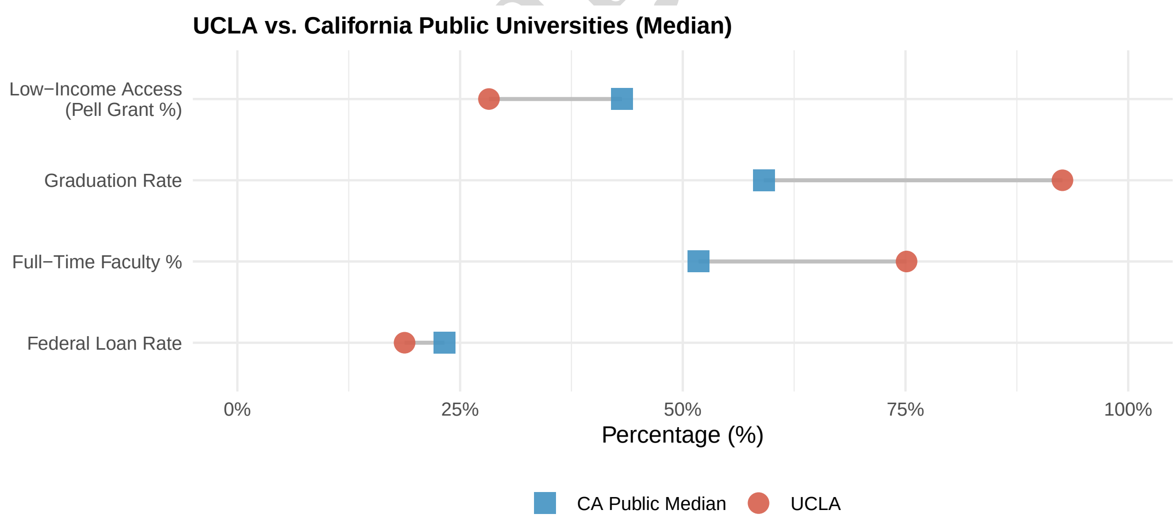
```
ca_pub <- full_data %>%
  filter(STABBR == "CA" & SCORECARD_SECTOR == 4 &
           INSTNM != "University of California-Los Angeles")

metrics_df <- data.frame(
  Metric     = c("Graduation Rate",
                "Low-Income Access\n(Pell Grant %)",
                "Federal Loan Rate",
                "Full-Time Faculty %"),
  UCLA       = c(ucla_data$C150_4    * 100,
                ucla_data$PCTPELL   * 100,
                ucla_data$PCTFLOAN * 100,
                ucla_data$PFTFAC    * 100),
  CA_Median = c(median(ca_pub$C150_4,   na.rm = TRUE) * 100,
                median(ca_pub$PCTPELL,  na.rm = TRUE) * 100,
                median(ca_pub$PCTFLOAN, na.rm = TRUE) * 100,
                median(ca_pub$PFTFAC,   na.rm = TRUE) * 100)
)

plot_long <- rbind(
```

```
  data.frame(Metric = metrics_df$Metric, Group = "UCLA",
             Value  = metrics_df$UCLA),
  data.frame(Metric = metrics_df$Metric, Group = "CA Public Median",
             Value  = metrics_df$CA_Median)
)

ggplot(plot_long, aes(x = Value, y = Metric, color = Group, shape = Group)) +
  geom_line(aes(group = Metric), color = "gray75", linewidth = 0.8) +
  geom_point(size = 4, alpha = 0.9) +
  scale_color_manual(
    values = c("UCLA" = "#D6604D", "CA Public Median" = "#4393C3")) +
  scale_shape_manual(
    values = c("UCLA" = 16, "CA Public Median" = 15)) +
  scale_x_continuous(
    labels = function(x) paste0(round(x, 0), "%"),
    limits = c(0, 100)) +
  labs(
    title = "UCLA vs. California Public Universities (Median)",
    x     = "Percentage (%)", y = NULL,
    color = NULL, shape = NULL
  ) +
  theme_minimal(base_size = 10) +
  theme(plot.title      = element_text(face = "bold", size = 10),
        legend.position = "bottom")
```



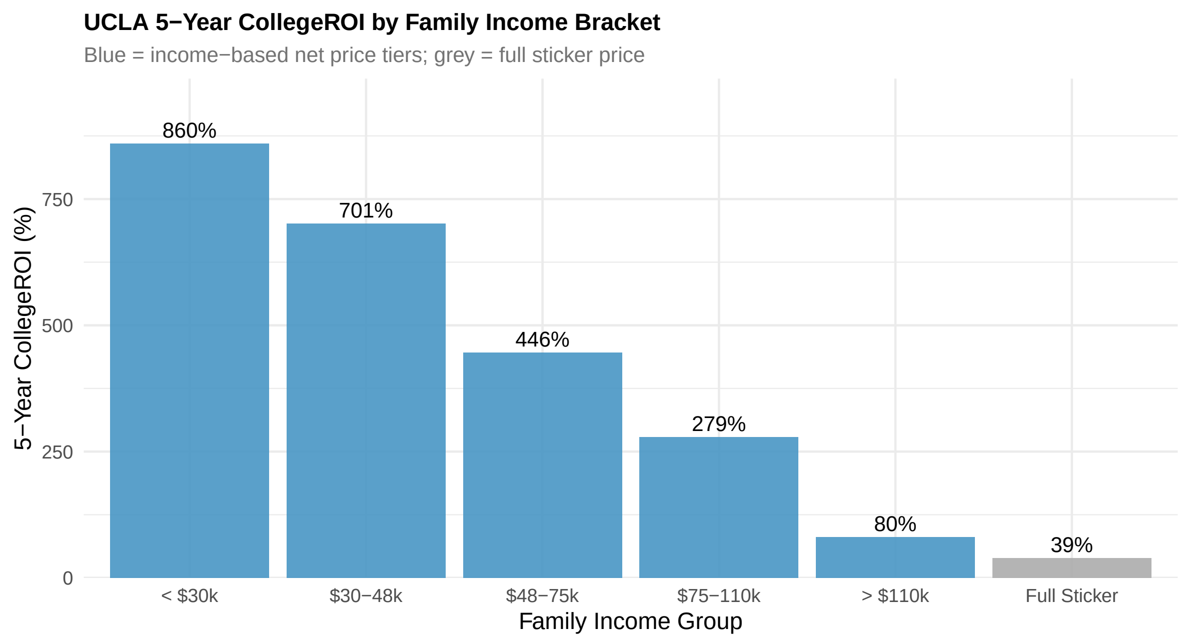
3.2 ROI by Family Income Bracket

UCLA is a public university with income-based net pricing. The sticker price (`COSTT4_A` = \$38,614/year) represents the full cost before any aid. The dataset includes five net-price tiers by family income (`NPT41_PUB` through `NPT45_PUB`), each reflecting what students in that income range actually paid on average. The ROI formula is applied to each tier using the same `wage_premium_5yr` computed above.

```
npt_df <- data.frame(
  Income_Group = factor(
    c("< $30k", "$30-48k", "$48-75k", "$75-110k", "> $110k", "Full Sticker"),
    levels = c("< $30k", "$30-48k", "$48-75k", "$75-110k", "> $110k", "Full
     ↪  Sticker")
  ),
  Annual_Cost = c(
    ucla_data$NPT41_PUB, ucla_data$NPT42_PUB, ucla_data$NPT43_PUB,
    ucla_data$NPT44_PUB, ucla_data$NPT45_PUB, ucla_data$COSTT4_A
  ),
  Is_Sticker = c(FALSE, FALSE, FALSE, FALSE, FALSE, TRUE)
) %>%
  mutate(
    Total_4yr = Annual_Cost * 4,
    ROI       = (5 * ucla_data$wage_premium_5yr - Total_4yr) / Total_4yr *
     ↪  100
  )

ggplot(npt_df, aes(x = Income_Group, y = ROI, fill = Is_Sticker)) +
  geom_col(alpha = 0.88) +
  geom_text(aes(label = paste0(round(ROI, 0), "%")),
            vjust = -0.4, size = 3.2) +
  scale_fill_manual(
    values = c("FALSE" = "#4393C3", "TRUE" = "#AAAAAA"),
    guide  = "none") +
  scale_y_continuous(expand = expansion(mult = c(0, 0.15))) +
  labs(
    title    = "UCLA 5-Year CollegeROI by Family Income Bracket",
    subtitle = "Blue = income-based net price tiers; grey = full sticker
     ↪  price",
    x        = "Family Income Group",
    y        = "5-Year CollegeROI (%)"
  ) +
  theme_minimal(base_size = 10) +
  theme(plot.title    = element_text(face = "bold", size = 10),
```

```
    plot.subtitle = element_text(color = "gray45", size = 8.5))
```



3.3 Earnings Growth Trajectory

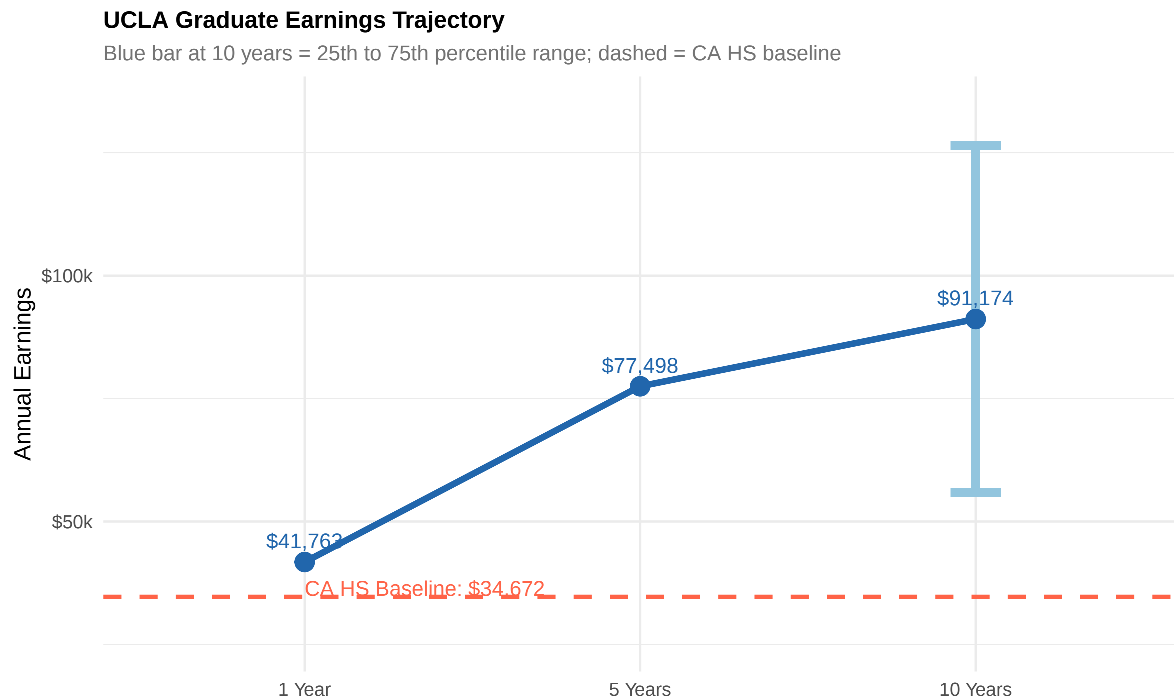
The chart below traces UCLA graduates' median earnings from one year to ten years post-entry (approximately age 30), relative to the California high-school earnings baseline. The error bar at the 10-year mark shows the 25th to 75th percentile range, illustrating how widely outcomes diverge over time.

```
traj_df <- data.frame(
  Period   = factor(c("1 Year", "5 Years", "10 Years"),
                    levels = c("1 Year", "5 Years", "10 Years")),
  Earnings = c(ucla_data$MD_EARN_WNE_1YR,
               ucla_data$MD_EARN_WNE_5YR,
               ucla_data$earn_10yr_mid)
)

range_df <- data.frame(
  Period = factor("10 Years", levels = c("1 Year", "5 Years", "10 Years")),
  ymin   = ucla_data$PCT25_EARN_WNE_P10,
  ymax   = ucla_data$PCT75_EARN_WNE_P10
)
```

```
baseline <- ucla_data$EARN_THR_STATE

ggplot(traj_df, aes(x = Period, y = Earnings, group = 1)) +
  geom_hline(yintercept = baseline, linetype = "dashed",
             color = "tomato", linewidth = 0.9) +
  geom_errorbar(data = range_df,
                aes(x = Period, ymin = ymin, ymax = ymax),
                inherit.aes = FALSE,
                color = "#92C5DE", width = 0.15, linewidth = 1.8) +
  geom_line(color = "#2166AC", linewidth = 1.3) +
  geom_point(color = "#2166AC", size = 3.5) +
  geom_text(aes(label = paste0("$", format(round(Earnings), big.mark =
   ↪  ","))),
            vjust = -0.9, size = 3.2, color = "#2166AC") +
  annotate("text",
           x     = 1,
           y     = baseline + 1800,
           label = paste0("CA HS Baseline: $",
                          format(round(baseline), big.mark = ",")),
           hjust = 0, color = "tomato", size = 3) +
  scale_y_continuous(
    labels = function(x) paste0("$", round(x / 1000), "k"),
    limits = c(25000, 135000)) +
  scale_x_discrete(limits = c("1 Year", "5 Years", "10 Years")) +
  labs(
    title    = "UCLA Graduate Earnings Trajectory",
    subtitle = "Blue bar at 10 years = 25th to 75th percentile range; dashed
     ↪  = CA HS baseline",
    x        = NULL,
    y        = "Annual Earnings"
  ) +
  theme_minimal(base_size = 10) +
  theme(plot.title    = element_text(face = "bold", size = 10),
        plot.subtitle = element_text(color = "gray45", size = 8.5))
```



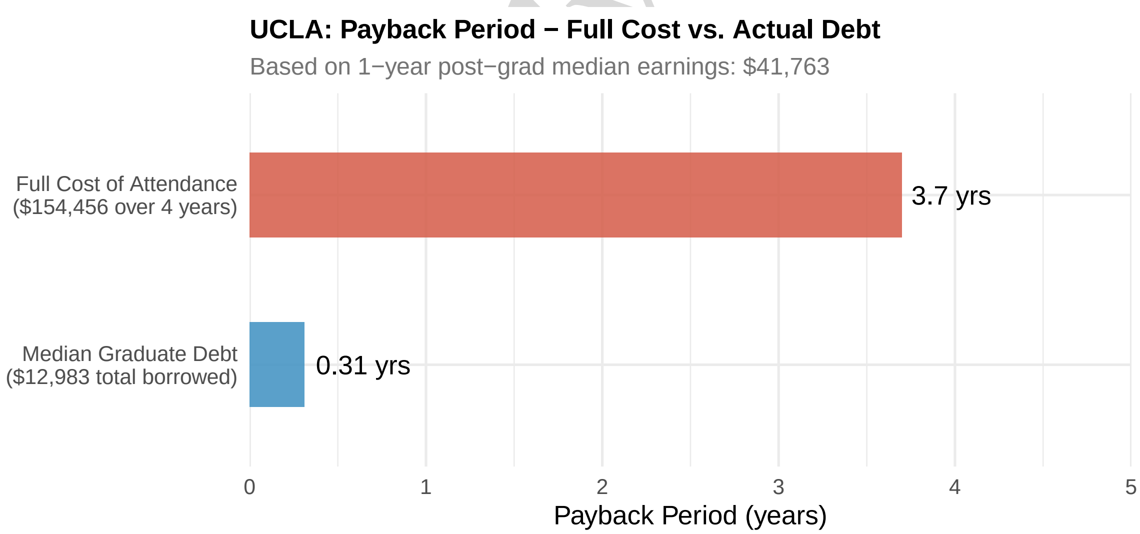
3.4 Full Cost vs. Actual Debt Burden

The standard payback period uses the full cost of attendance as the investment base. However, the median **borrower** at UCLA had only $12,983 in total federal loans — a fraction of the $154,456 four-year sticker cost, reflecting that only 18.8% of students took out federal loans and many received grants or institutional aid. The chart compares how long it takes to earn back each amount using first-year post-graduation earnings.

```
payback_compare <- data.frame(
  Scenario = c(
    paste0("Full Cost of Attendance\n($",
           format(round(ucla_data$total_cost_4yr), big.mark = ","),
           " over 4 years)"),
    paste0("Median Graduate Debt\n($",
           format(round(ucla_data$DEBT_MDN), big.mark = ","),
           " total borrowed)")
  ),
  Payback_yrs = c(
    ucla_data$total_cost_4yr / ucla_data$MD_EARN_WNE_1YR,
    ucla_data$DEBT_MDN       / ucla_data$MD_EARN_WNE_1YR
  )
```

```
)

ggplot(payback_compare,
       aes(x = reorder(Scenario, Payback_yrs), y = Payback_yrs,
           fill = reorder(Scenario, Payback_yrs))) +
  geom_col(alpha = 0.88, width = 0.5) +
  geom_text(aes(label = paste0(round(Payback_yrs, 2), " yrs")),
            hjust = -0.12, size = 3.5) +
  coord_flip() +
  scale_fill_manual(values = c("#4393C3", "#D6604D"), guide = "none") +
  scale_y_continuous(limits = c(0, 5), expand = c(0, 0)) +
  labs(
    title    = "UCLA: Payback Period - Full Cost vs. Actual Debt",
    subtitle = paste0("Based on 1-year post-grad median earnings: $",
                      format(round(ucla_data$MD_EARN_WNE_1YR), big.mark =
                       ↪  ",")),
    x        = NULL,
    y        = "Payback Period (years)"
  ) +
  theme_minimal(base_size = 10) +
  theme(plot.title    = element_text(face = "bold", size = 10),
        plot.subtitle = element_text(color = "gray45", size = 8.5))
```



3.5 Risk-Return Profile

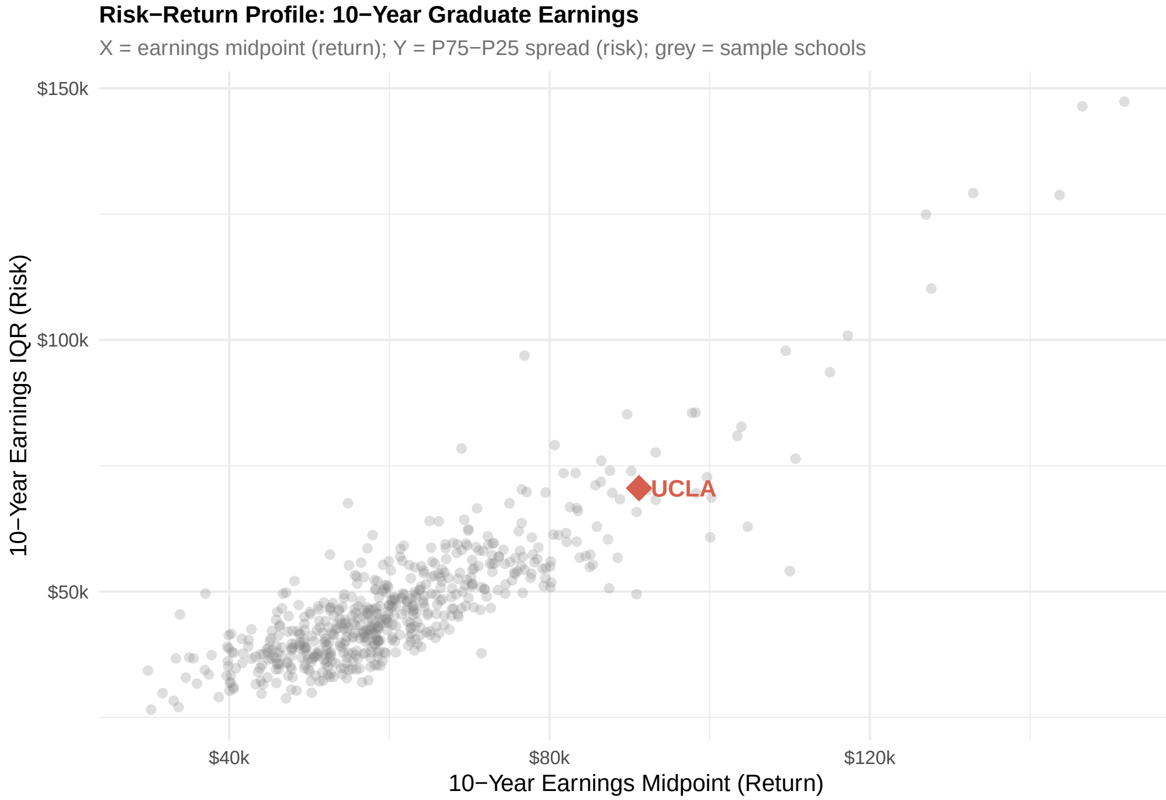
Not all schools with similar median earnings are equally predictable. The scatter plot below positions each school in the sample by its 10-year earnings midpoint (the "return") against the spread between its 75th and 25th percentile earnings (the "risk" — a wider spread means more variable graduate outcomes).

UCLA is highlighted in red.

```
rr_df <- sample_data %>%
  filter(!is.na(PCT25_EARN_WNE_P10) & !is.na(PCT75_EARN_WNE_P10) &
           !is.na(earn_10yr_mid)) %>%
  mutate(earnings_iqr = PCT75_EARN_WNE_P10 - PCT25_EARN_WNE_P10)

ucla_rr <- data.frame(
  earn_10yr_mid = ucla_data$earn_10yr_mid,
  earnings_iqr  = ucla_data$PCT75_EARN_WNE_P10 - ucla_data$PCT25_EARN_WNE_P10
)

ggplot(rr_df, aes(x = earn_10yr_mid, y = earnings_iqr)) +
  geom_point(alpha = 0.25, color = "gray50", size = 1.5) +
  geom_point(data = ucla_rr, aes(x = earn_10yr_mid, y = earnings_iqr),
             color = "#D6604D", size = 5, shape = 18) +
  annotate("text",
           x     = ucla_rr$earn_10yr_mid + 1500,
           y     = ucla_rr$earnings_iqr,
           label = "UCLA",
           hjust = 0, color = "#D6604D", fontface = "bold", size = 3.5) +
  scale_x_continuous(labels = function(x) paste0("$", round(x / 1000), "k"))
   ↪ +
  scale_y_continuous(labels = function(x) paste0("$", round(x / 1000), "k"))
   ↪ +
  labs(
    title    = "Risk-Return Profile: 10-Year Graduate Earnings",
    subtitle = "X = earnings midpoint (return); Y = P75-P25 spread (risk);
     ↪ grey = sample schools",
    x        = "10-Year Earnings Midpoint (Return)",
    y        = "10-Year Earnings IQR (Risk)"
  ) +
  theme_minimal(base_size = 10) +
  theme(plot.title    = element_text(face = "bold", size = 10),
        plot.subtitle = element_text(color = "gray45", size = 8.5))
```



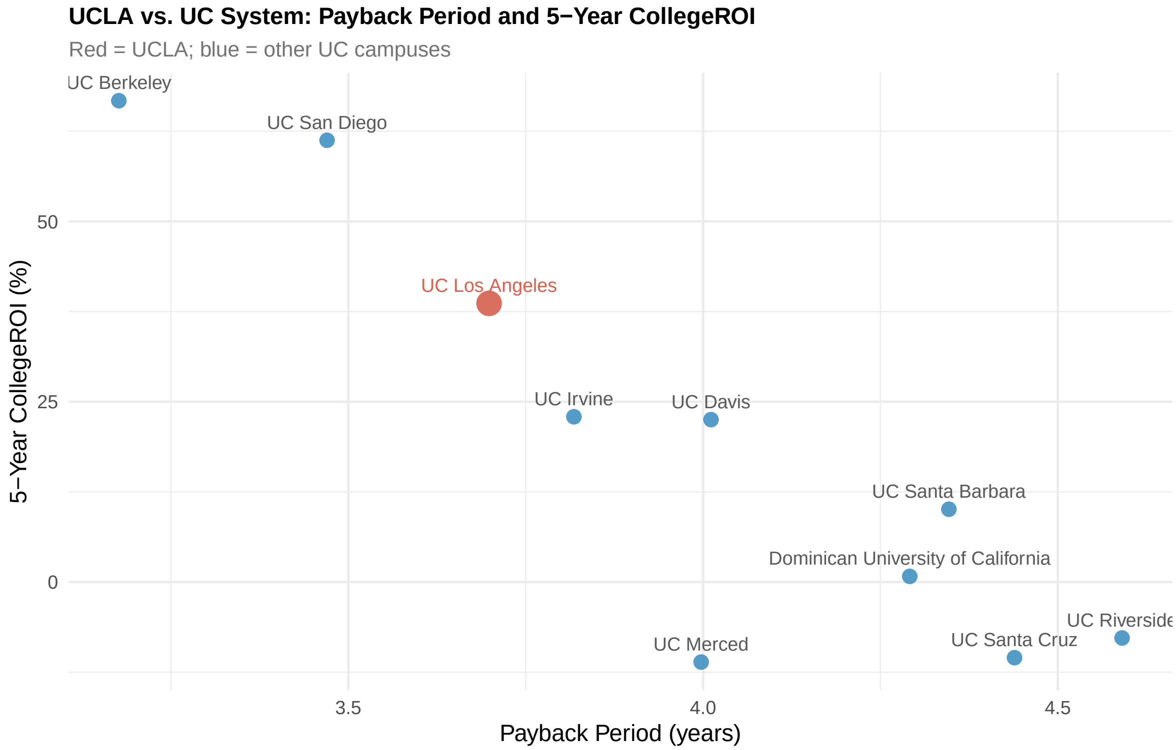
3.6 UCLA vs. the UC System

Finally, UCLA is compared to other University of California campuses on the two primary financial metrics: payback period and 5-year CollegeROI. This comparison is particularly meaningful since all UC schools share the same state baseline earnings, the same in-state tuition structure, and operate within the same California labor market.

```
uc_schools <- full_data %>%
  filter(grepl("University of California", INSTNM)) %>%
  mutate(
    total_cost_4yr    = 4 * COSTT4_A,
    payback_period    = total_cost_4yr / MD_EARN_WNE_1YR,
    wage_premium_5yr = MD_EARN_WNE_5YR - EARN_THR_STATE,
    college_roi       = (5 * wage_premium_5yr - total_cost_4yr) /
                         total_cost_4yr * 100,
    is_ucla           = INSTNM == "University of California-Los Angeles",
    label             = gsub("University of California-", "UC ", INSTNM)
  ) %>%
```

```r
  filter(!is.na(college_roi) & !is.na(payback_period) & payback_period > 0)

ggplot(uc_schools,
       aes(x = payback_period, y = college_roi,
           color = is_ucla, size = is_ucla)) +
  geom_point(alpha = 0.9) +
  geom_text(aes(label = label),
            vjust = -0.85, size = 2.9,
            color = ifelse(uc_schools$is_ucla, "#D6604D", "gray35")) +
  scale_color_manual(values = c("TRUE" = "#D6604D", "FALSE" = "#4393C3"),
                     guide  = "none") +
  scale_size_manual(values  = c("TRUE" = 4.5, "FALSE" = 2.5),
                    guide   = "none") +
  labs(
    title    = "UCLA vs. UC System: Payback Period and 5-Year CollegeROI",
    subtitle = "Red = UCLA; blue = other UC campuses",
    x        = "Payback Period (years)",
    y        = "5-Year CollegeROI (%)"
  ) +
  theme_minimal(base_size = 10) +
  theme(plot.title    = element_text(face = "bold", size = 10),
        plot.subtitle = element_text(color = "gray45", size = 8.5))
```


UCLA vs. UC System: Payback Period and 5−Year CollegeROI
Red = UCLA; blue = other UC campuses
UC Berkeley
UC San Diego
UC Los Angeles
UC Irvine
UC Davis
UC Santa Barbara
Dominican University of California
UC Merced
UC Santa Cruz
UC Riverside
50
25
0
3.5
4.0
4.5
5−Year CollegeROI (%)
Payback Period (years)