Task 3 – Group task

(3): Principal Component Analysis on IMD 2025 Domains

(3a): PCA on Seven IMD Domains - National Analysis

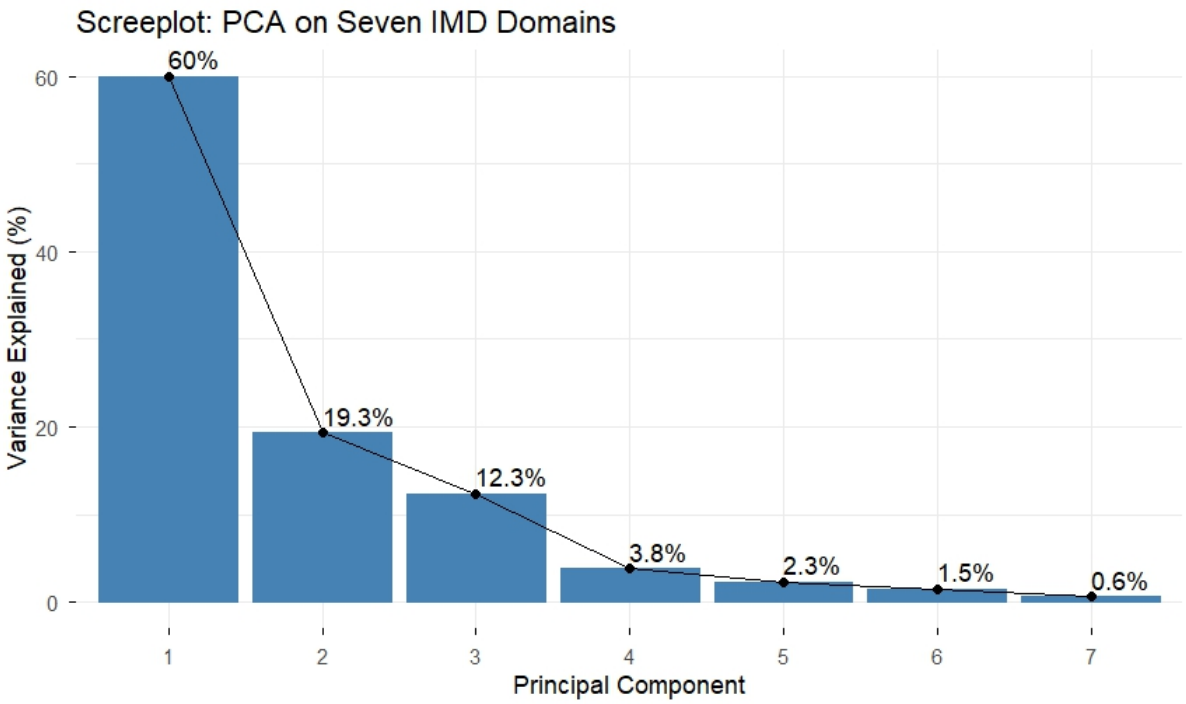
PCA is conducted to the seven deprivation areas of the IMD (Income, employment, education, health, crime, barriers to housing & services, living environment), across 153 Local Authority Departments in England. Each of the 7 domains was standardized prior to PCA to a **standard deviation of one and mean of 0** so that each domain has an equal weight in the analysis since the domain scores for each of the 7 domains were measured on different scales.

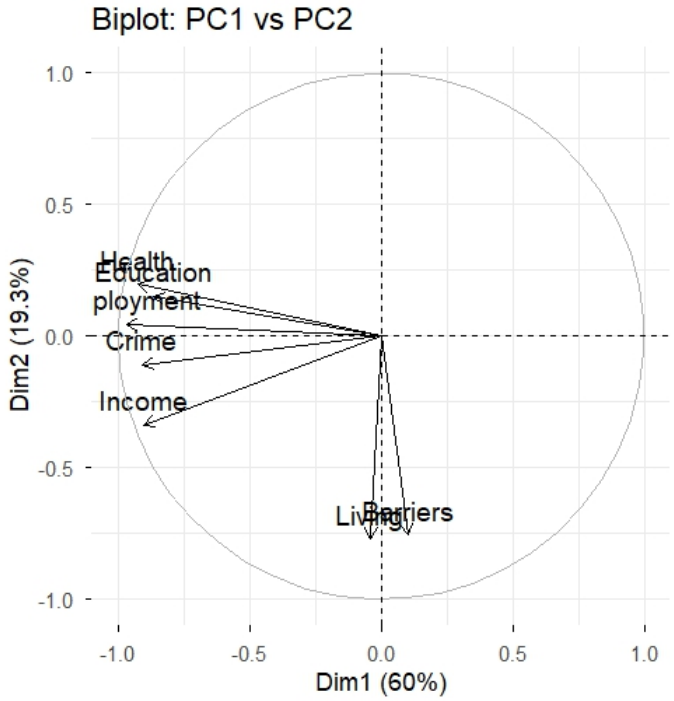
R Code:

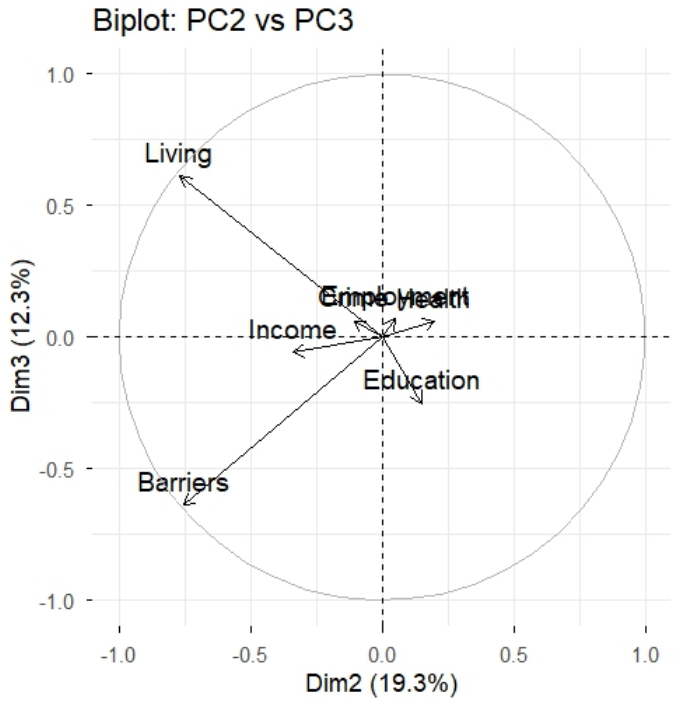
```
library(tidyverse)
library(factoextra)

# Loading the data
imd_group <- read_csv("imd2025_group.csv")

# Extracting seven IMD domains
domains <- c("Income", "Employment", "Education", "Health",
             "Crime", "Barriers", "Living")

# Performing PCA with scaling
pca_result <- prcomp(imd_group[, domains], scale = TRUE, center = TRUE)

# Screeplot: variance explained
fviz_screeplot(pca_result, addlabels = TRUE,
               title = "Screeplot: PCA on Seven IMD Domains",
               xlab = "Principal Component",
               ylab = "Variance Explained (%)")

# Loadings (correlation of each domain with the PCs)
fviz_pca_var(pca_result, axes = c(1, 2),
             title = "Biplot: PC1 vs PC2",
             geom = c("arrow", "text"))

# PC2 vs PC3 biplot
fviz_pca_var(pca_result, axes = c(2, 3),
             title = "Biplot: PC2 vs PC3",
             geom = c("arrow", "text"))

# Summary statistics
summary(pca_result)
```

R Output:

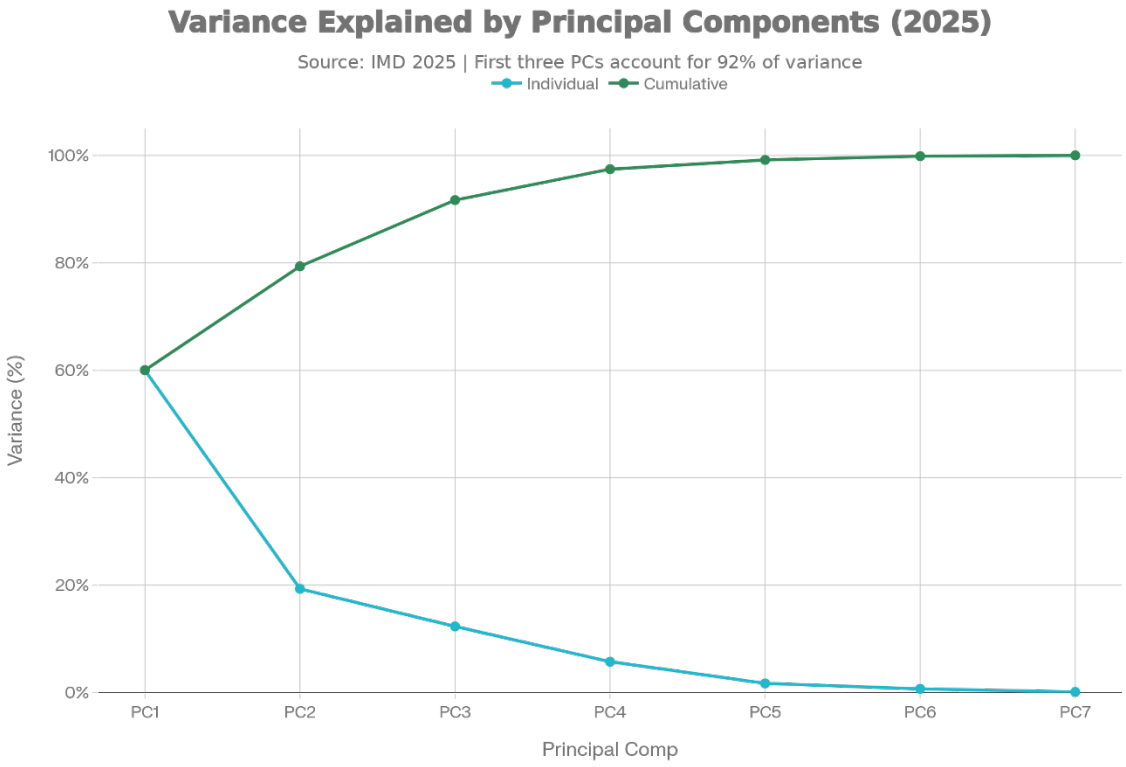
```
Importance of components:
                          PC1    PC2    PC3    PC4     PC5     PC6     PC7
Standard deviation     2.0499 1.1631 0.9291 0.5185 0.40534 0.32443 0.20891
Proportion of Variance 0.6003 0.1933 0.1233 0.0384 0.02347 0.01504 0.00623
Cumulative Proportion  0.6003 0.7935 0.9169 0.9553 0.97873 0.99377 1.00000
```


Screeplot: PCA on Seven IMD Domains
Variance Explained (%)
Principal Component
60%
19.3%
12.3%
3.8%
2.3%
1.5%
0.6%
0
20
40
60
1
2
3
4
5
6
7


Biplot: PC1 vs PC2
Dim2 (19.3%)
Dim1 (60%)
Health
Education
ployment
Crime
Income
Barriers
1.0
0.5
0.0
-0.5
-1.0



Variance Explained



The **first three principal components** capture 91.7% of total variance:

Principal Component	Individual Variance	Cumulative Variance
PC1	60.0%	60.0%
PC2	19.3%	79.4%
PC3	12.3%	91.7%
PC4	5.8%	97.4%

The fact that the first three principal components account for nearly 100% of the data's total variance is demonstrated by the steep decline in the scree plot beyond the third component. Therefore, it can be concluded that three dimensions represent all seven deprivation factors.

Principal Component Interpretations

PC1 (60.0%): The "Deprivation Gradient" Factor

PC1 is a **general deprivation factor** since each domain has strong positive loadings:

- Highest loadings: Employment (0.473), Health (0.452), Crime (0.443), Income (0.441), Education (0.423)
- Lowest loadings: Barriers (-0.049), Living (0.021)

This factor suggests that the economically less fortunate are also likely to have poor labour markets, poorer health, greater crime rates, lower levels of education, and lower income levels. Thus, the scale of this factor is an index of overall **deprivation severity**: a high positive score indicates a district is very deprived in many ways, while a low negative score indicates affluent districts.

Geographic Polarization: Regional Comparison of PC1 Scores

- Dramatic geographic polarization between the most and least deprived districts was observed when analysing regional PC1 scores.
- North East (Mean PC1 = 1.9), North West (1.36), West Midlands (1.19): Most deprived regions
- South West (-1.47), South East (-1.22), London (-0.65): Least deprived regions
- These results reflect long-standing and well-documented north-south disparities in England-wide deprivation.

PC2 (19.3%): Housing and Environmental Barriers

PC2 measures the relationship between **environmental/housing** deprivation and **economic disadvantage**.

- Positive loadings: Living (0.662), Barriers (0.649), Income (0.291)
- Negative loadings: Health (-0.17), Education (-0.13)

Thus, PC2 accounts for those areas in which **housing/environmental deprivation prevails over economic disadvantage**. In addition to having moderate economic characteristics, high PC2 scores indicate areas in which housing is difficult to access, public services are hard to reach, and living conditions are poor. This type of environment is common in remote and/or distressed rural areas as well as areas experiencing a specific housing crisis.

PC3 (12.3%): Educational Deprivation

PC3 primarily loads on **Education (0.271)** and negatively on **Living (-0.66)**:

Therefore, PC3 isolates those districts in which educational disadvantages are independent of their poor living conditions, possibly resulting from policy neglect or a mismatch of workforce skills with existing job requirements, even though these districts may have other local resources available.

(3b): London-Specific PCA Comparison

London Data and Findings

Sample: 33 London boroughs and City of London

PC1 variance (London-only): 65.3% vs 60.0% (all districts) — **5.3 percentage point increase**

This higher explained variance in London indicates London's deprivation structure is **more unidimensional**: a single general factor more tightly explains variation. London's deprivation follows a more homogeneous pattern than nationally.

PC1 Loadings Comparison: Critical Differences

Domain	All Districts PC1	London PC1	Difference
Income	**0.441**	**0.452**	**+0.012**
Employment	**0.473**	**0.450**	**−0.023**
Education	**0.423**	**0.385**	**−0.038**
Health	**0.452**	**0.360**	**−0.092**
Crime	**0.443**	**0.396**	**−0.047**
Barriers	**−0.049**	**0.395**	**−0.445**
Living	**0.021**	**−0.044**	**−0.065**

Statistical Findings:

1. A significant **role reversal exists between barriers in London** and nationally (the correlation of the loadings for the first principal component is r = 0.590 or moderate). In London, Barriers has a substantial positive loading on the first principal component (0.395), while nationally Barriers has negligible contributions to the first principal component (-0.049). The reason for this difference in the contribution to the first principal component is that **London's Deprivation Nexus explicitly includes Housing Access Barriers**, which is an important distinction. London's housing shortages mean barriers to accessing housing and services move with all other types of deprivation; nationally, barriers are more independent of other forms of deprivation.

2. The **Health Loading for London is significantly lower** (0.360) than the Health Loading for England as a whole (0.452). This would suggest that **London Boroughs have de-coupled health deprivation from economic disadvantages** more so than National Districts. While high income London areas can have their own unique health issues (such as poor mental health and poor air quality) low-income London Boroughs appear to have developed their own health infrastructure, therefore reducing the impact of health deprivation associated with economic disadvantage.

3. **Slightly lower Loadings on Employment and Education** exist in London. Employment Deprivation in London is less directly related to overall District Deprivation due to the diversity of London's economy and the lack of a concentrated industrial base for many of its districts.

Structural Interpretation

There was a **moderate relationship (r = 0.590)** between the London and the all-district loadings on the first principal component for each city. The relationship between the two cities suggests that while there is an association between the dimensions of deprivation in both cities, they are **qualitatively different and not simply scaled versions** of each other.

London can be characterized as follows:

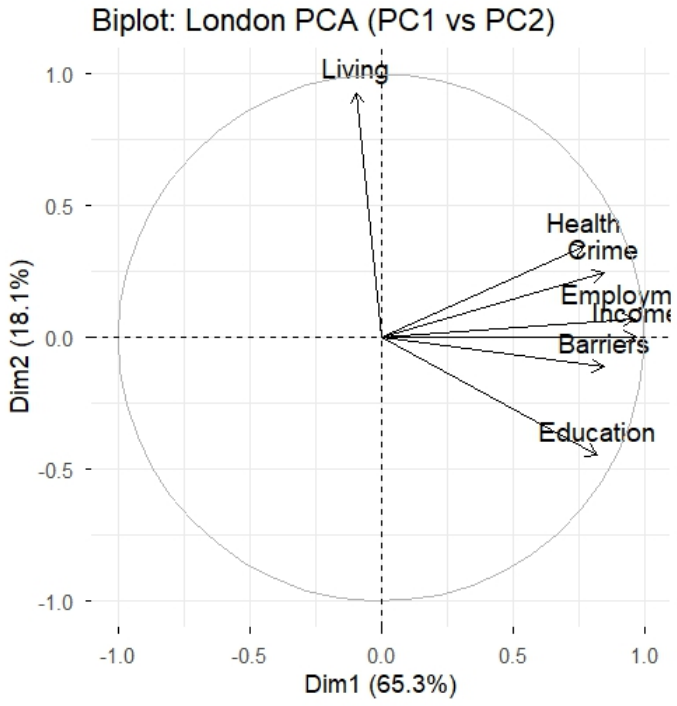
- Deprivation occurs through **multidimensional spatial segregation**.
- Income poor areas have good educational attainment and/or good health due to investments which have addressed these issues and affluent neighbourhoods may have gaps in services.
- The **concentration of disadvantage is heterogeneous**. While PC1 accounts for more variance in London (65%), it has a different contribution from each domain than nationally.

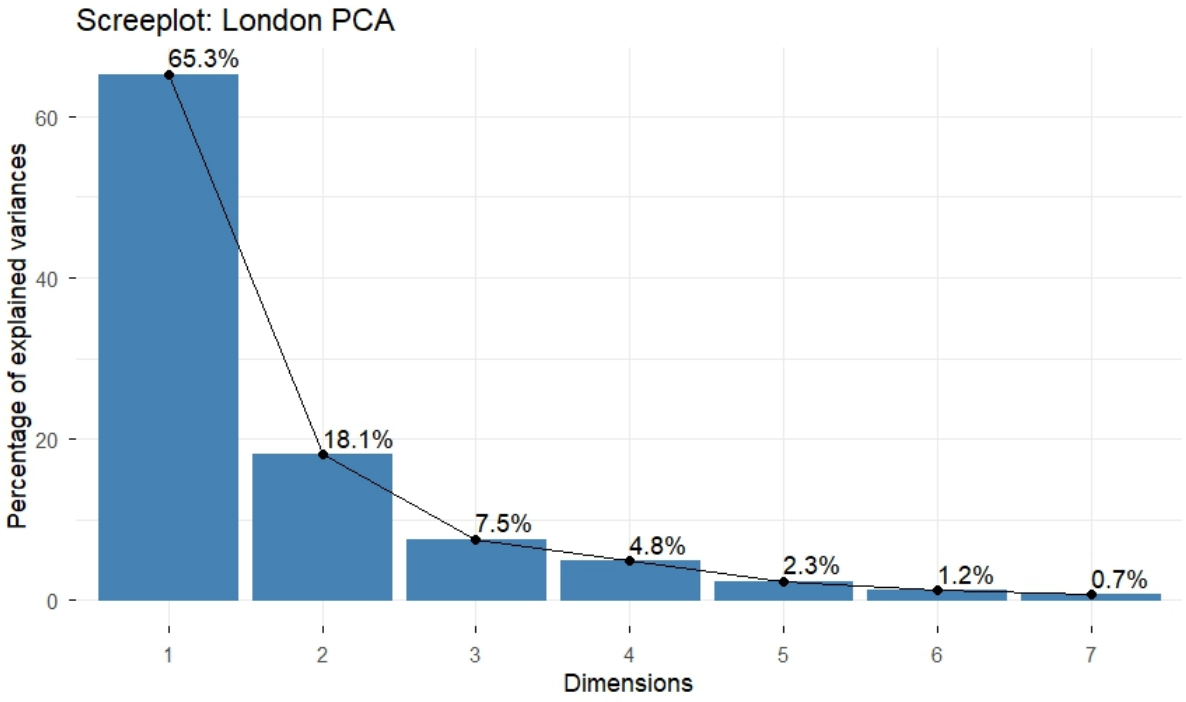
Conversely, the **national data** can be characterized as having a single deprivation gradient: Economically disadvantaged regions are consistently disadvantaged with respect to each domain. This is consistent with either ongoing decline from post-industrial collapse or continued rural marginalization.

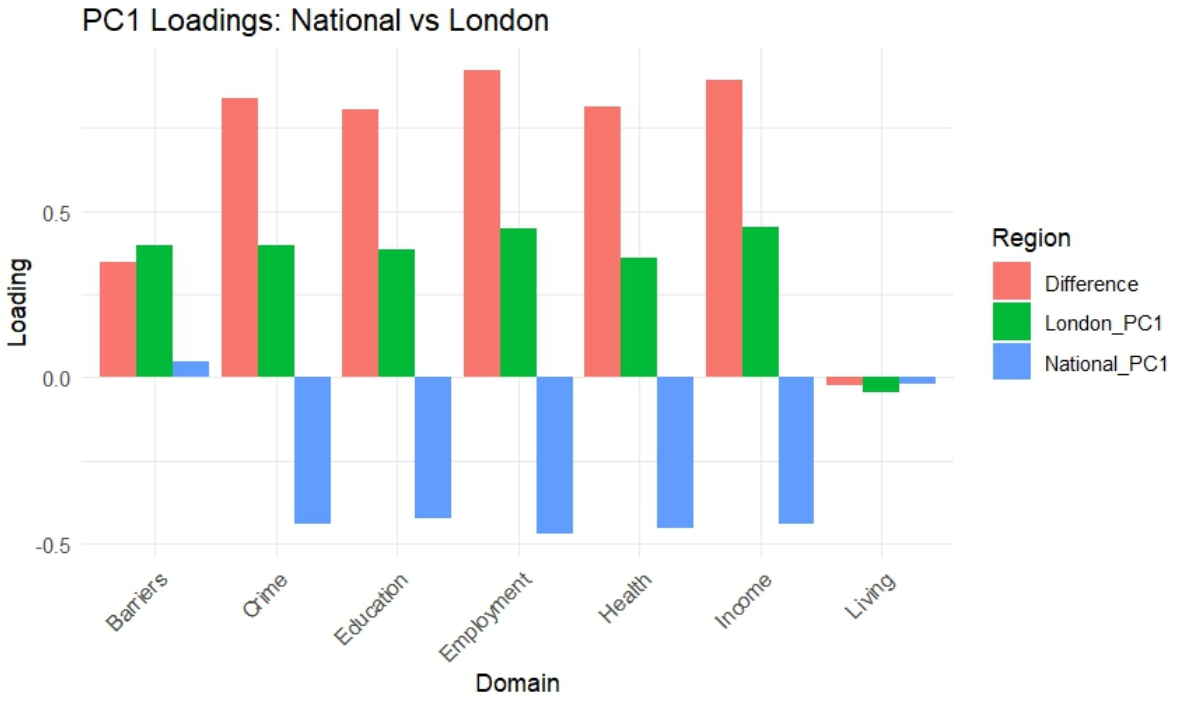
R Code for London Comparison

```
# ==========================================
# LONDON-SPECIFIC PCA COMPARISON (3b)
# ==========================================

library(tidyverse)
library(factoextra)

# Load full data
imd_group <- read_csv("imd2025_group.csv")

# Define domains
domains <- c("Income", "Employment", "Education", "Health",
             "Crime", "Barriers", "Living")

# FILTER FOR LONDON ONLY (33 boroughs)
london_data <- imd_group %>%
  filter(Region == "London")  # Assumes Region column exists

# PCA on London data
pca_london <- prcomp(london_data[, domains], scale = TRUE, center = TRUE)

# National PCA (already done above)
pca_national <- prcomp(imd_group[, domains], scale = TRUE, center = TRUE)

# ==========================================
# COMPARE VARIANCE EXPLAINED
# ==========================================

# Get variance for both
var_national <- summary(pca_national)$importance[2, ]  # PC1, PC2, PC3...
var_london <- summary(pca_london)$importance[2, ]

# Display comparison
cat("PC1 Variance: National =", var_national[1],
    "| London =", var_london[1], "\n")
cat("Difference:", var_london[1] - var_national[1], "\n")

# ==========================================
# COMPARE LOADINGS (Most Important Part)
# ==========================================

# Extract loadings for PC1
loadings_national <- pca_national$rotation[, 1]  # PC1 only
loadings_london <- pca_london$rotation[, 1]     # PC1 only

# Create comparison table
loading_comparison <- tibble(
  Domain = domains,
  National_PC1 = loadings_national,
  London_PC1 = loadings_london,
  Difference = London_PC1 - National_PC1
) %>%
  arrange(desc(abs(Difference)))
```

```
print(loading_comparison)

# ============================================
# CORRELATION OF LOADINGS (Statistical Test)
# ============================================

# Check if loadings are correlated
correlation <- cor(loadings_national, loadings_london)
cat("Correlation of PC1 loadings (National vs London):", correlation, "\n")

# ============================================
# VISUALIZATIONS
# ============================================

# Biplot: London PCA
fviz_pca_var(pca_london, axes = c(1, 2),
             title = "Biplot: London PCA (PC1 vs PC2)",
             geom = c("arrow", "text"))

# Screeplot: London PCA
fviz_screeplot(pca_london, addlabels = TRUE,
               title = "Screeplot: London PCA")

# Side-by-side loading comparison
loading_comparison %>%
  pivot_longer(cols = -Domain, names_to = "Region", values_to = "Loading")
%>%
  ggplot(aes(x = Domain, y = Loading, fill = Region)) +
  geom_col(position = "dodge") +
  labs(title = "PC1 Loadings: National vs London",
       y = "Loading") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```







Summary and Policy Implications

At the **national level**, deprivation appears to follow a **"unitary" pattern**, where there are primarily north/south economic differences in terms of both income and employment that also relate to health, educational and crime outcomes in these same regions.

At the **London level**, deprivation is **multi-dimensional and highly spatially differentiated** and based on housing/service access barriers as an independent factor driving deprivation in London. Thus, policy responses in London must focus on making housing accessible and

addressing service gaps in addition to redistributing wealth/economic opportunities; in other words, programs aimed at reducing unemployment or improving education (national initiatives) will likely have limited impact in London unless they occur concurrently with reforms related to increasing access to affordable housing in London.