

# ■ R Exam Day Toolkit — Search■to■Solve Edition

## 1. What To Do FIRST (Every Exam Question)

- 1) Load libraries:  
library(tidyverse)  
library(GGally)
- 2) Inspect data:  
head(df)  
str(df)  
summary(df)
- 3) Identify question type (plot, summarise, model, reshape).

## 2. The 10 Google Searches That Solve 95% of Exam Questions

Histogram:  
"R ggplot histogram geom\_histogram geom\_vline example"

Scatterplot:  
"R ggplot scatterplot geom\_point geom\_smooth example"

Group summaries:  
"R dplyr group\_by summarise median example"

Counting categories:  
"R dplyr count n\_distinct example"

Missing values:  
"R dplyr filter NA example"

Pivoting data:  
"R tidyr pivot\_longer example"  
"R tidyr pivot\_wider example"

Join datasets:  
"R dplyr left\_join example"

Regression:  
"R lm polynomial interaction example"  
"R AIC compare models example"

## 3. Most Useful Code Patterns (Copy, Replace Columns)

FILTER ROWS:  
df\_clean <- df %>% filter(!is.na(col1), !is.na(col2))

COUNT:  
df %>% count(country, sort = TRUE)  
n\_distinct(df\$country)

GROUP SUMMARY:

```

df %>% group_by(sector) %>% summarise(med_gap = median(Gap, na.rm=TRUE))

HISTOGRAM:
ggplot(df, aes(Gap)) + geom_histogram(binwidth = 2)

HISTOGRAM + PERCENTILES:
p05 <- quantile(df$Gap, 0.05)
p95 <- quantile(df$Gap, 0.95)
ggplot(df, aes(Gap)) +
  geom_histogram() +
  geom_vline(xintercept=p05) +
  geom_vline(xintercept=p95)

SCATTERPLOT:
ggplot(df, aes(X, Y)) + geom_point() + geom_smooth(method="lm")

REORDER FACTORS:
df$country <- factor(df$country, levels = df %>% group_by(country) %>% summarise(med=median(Y))

PIVOT LONG:
df_long <- df %>% pivot_longer(cols = starts_with("sector"), names_to="sector", values_to="valu

JOIN:
joined <- left_join(df1, df2, by="id")

```

## 4. Regression Code Patterns

```

FULL MODEL TEMPLATE:
mod_full <- lm(Y ~ poly(X1,2) + poly(X2,2) + X3 + X4*Factor, data=df)

REMOVE TERM:
mod2 <- update(mod_full, . ~ . - X1)

CHECK MODELS:
AIC(mod_full, mod2)

DIAGNOSTICS:
par(mfrow=c(2,2)); plot(mod2); par(mfrow=c(1,1))

VIF:
library(car); vif(mod2)

```

## 5. One-Line Interpretation Templates

```

Histogram:
"Distribution is right-skewed with most values between X and Y."

Scatterplot:
"There is a positive relationship; as X increases, Y tends to increase."

Group summary:
"Sector A has the highest median, Sector B the lowest."

Regression:
"X1 increases Y; X2 has a nonlinear effect; interaction indicates effect depends on category."

```

Diagnostics:

"Residuals show no major violations of linearity or constant variance."

## 6. Exam Decision Tree

Q1: Is it about distribution? → Make histogram + quantiles.

Q2: Two variables? → Scatterplot + smooth line.

Q3: Compare groups? → group\_by + summarise + reorder + boxplot.

Q4: Many similar columns? → pivot\_longer.

Q5: Need regression? → Full model → Remove terms → Compare AIC → Diagnostics.

Q6: Need differences between years? → pivot\_wider + mutate(diff = col2023 - col2022).

## 7. Emergency Troubleshooting

Plot not working? → aes() variable name has a space → wrap in backticks: `Mean Gap`.

Summaries failing? → Missing values → add na.rm=TRUE.

Factor ordering wrong? → Use factor(levels=...).

Join failing? → Column names differ → rename before join.

Histogram ugly? → Add binwidth.

Scatter messy? → alpha=0.5 to make points transparent.