

UNIVERSIDAD NACIONAL DEL ALTIPLANO  
FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA  
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E  
INFORMÁTICA

---

ANÁLISIS ESPACIAL DE  
REGIONES NATURALES EN EL  
DEPARTAMENTO DE PUNO

---

Encuesta Agrícola Nacional 2024

**Presentado por:**

Willy Vilca Apaza

**Carrera:** Ingeniería Estadística e Informática

**Código:** 214654

**Año Académico:** 2025

**Semestre:** X

Puno, Perú  
10 de setiembre de 2025

# Índice

<b>1. Introducción</b>	<b>2</b>
1.1. Contexto Geográfico . . . . .	2
1.2. Justificación . . . . .	2
1.3. Objetivos . . . . .	2
1.3.1. Objetivo General . . . . .	2
1.3.2. Objetivos Específicos . . . . .	3
<b>2. Metodología</b>	<b>3</b>
2.1. Fuente de Datos . . . . .	3
2.2. Herramientas Computacionales . . . . .	3
2.3. Variables del Dataset . . . . .	3
2.4. Proceso Metodológico . . . . .	4
2.5. Coordenadas Geográficas . . . . .	4
<b>3. Resultados</b>	<b>5</b>
3.1. Distribución Regional por Provincia . . . . .	5
3.2. Estadísticas Descriptivas . . . . .	5
3.3. Mapas Generados . . . . .	5
3.3.1. Mapa Estático . . . . .	5
3.3.2. Gráfico de Barras . . . . .	6
3.3.3. Mapa Interactivo . . . . .	6
3.4. Análisis por Región . . . . .	6
3.4.1. Provincias de Sierra . . . . .	6
3.4.2. Provincias de Selva . . . . .	7
<b>4. Discusión</b>	<b>7</b>
4.1. Coherencia con la Geografía Regional . . . . .	7
4.2. Implicaciones para Políticas Públicas . . . . .	7
4.3. Limitaciones del Estudio . . . . .	7
<b>5. Código R Utilizado</b>	<b>8</b>
<b>6. Conclusiones</b>	<b>8</b>
<b>7. Recomendaciones</b>	<b>9</b>
<b>8. Referencias</b>	<b>9</b>

# 1. Introducción

El presente documento desarrolla un análisis espacial exhaustivo de las regiones naturales en el departamento de Puno, utilizando datos de la Encuesta Agrícola Nacional 2024. El objetivo principal es visualizar y analizar la distribución geográfica de las tres regiones naturales del Perú (Costa, Sierra y Selva) en las 13 provincias que conforman este departamento altiplánico.

## 1.1. Contexto Geográfico

Puno es un departamento del sureste peruano que se caracteriza por su extraordinaria diversidad geográfica. Ubicado principalmente en la región andina, el departamento presenta características únicas que lo distinguen:

- **Altiplano:** La meseta del Collao, una de las más altas del mundo (3,800 - 4,000 msnm)
- **Lago Titicaca:** El lago navegable más alto del planeta
- **Cordillera Oriental:** Zona de transición hacia la Amazonía
- **Selva Alta:** Provincias del norte con características amazónicas

Esta diversidad geográfica genera patrones espaciales específicos que este análisis busca identificar y caracterizar mediante técnicas de visualización geoespacial.

## 1.2. Justificación

El análisis espacial de regiones naturales es fundamental por varias razones:

1. **Planificación agrícola:** Permite diseñar políticas diferenciadas según características regionales
2. **Ordenamiento territorial:** Facilita la comprensión de la distribución geográfica
3. **Desarrollo sostenible:** Identifica áreas con potencialidades específicas
4. **Investigación científica:** Genera conocimiento sobre patrones espaciales

## 1.3. Objetivos

### 1.3.1. Objetivo General

Realizar un análisis espacial de la distribución de regiones naturales en las 13 provincias del departamento de Puno mediante técnicas de geoestadística y visualización espacial.

1.3.2. Objetivos Específicos

- Procesar y limpiar los datos de la Encuesta Agrícola Nacional 2024
- Identificar la región natural predominante en cada provincia de Puno
- Generar mapas estáticos e interactivos de alta calidad
- Analizar estadísticamente los patrones de distribución regional
- Interpretar los resultados en el contexto geográfico del departamento

2. Metodología

2.1. Fuente de Datos

Se utilizó el dataset `21_CAP1200B_ME.sav` correspondiente a la Encuesta Agrícola Nacional 2024. Este archivo en formato SPSS contiene 49,289 observaciones a nivel nacional, de las cuales 4,951 corresponden al departamento de Puno.

2.2. Herramientas Computacionales

El análisis se desarrolló en R versión 4.5.1 utilizando las siguientes librerías especializadas:

Cuadro 1: Librerías utilizadas en el análisis espacial	
Librería	Función Principal
haven	Lectura de archivos SPSS (.sav)
sf	Análisis espacial (Simple Features)
ggplot2	Visualización gráfica avanzada
leaflet	Mapas interactivos web
dplyr	Manipulación de datos
viridis	Paletas de colores
htmlwidgets	Exportación de mapas HTML

2.3. Variables del Dataset

El dataset contiene 25 variables, de las cuales se utilizaron principalmente:

- **NOMBREDD**: Nombre del departamento
- **NOMBREPV**: Nombre de la provincia
- **REGION**: Código de región natural (1=Costa, 2=Sierra, 3=Selva)
- **CCDD, CCPP, CCDI**: Códigos geográficos

## 2.4. Proceso Metodológico

El análisis se estructuró en 12 fases secuenciales:

1. **Carga de datos:** Importación del archivo SPSS (49,289 obs.)
2. **Exploración inicial:** Identificación de 25 variables disponibles
3. **Filtrado geográfico:** Selección de datos de Puno (4,951 obs.)
4. **Identificación de provincias:** Validación de las 13 provincias
5. **Agregación:** Determinación de región predominante por provincia
6. **Asignación de coordenadas:** Incorporación de lat/lon geográficas
7. **Integración:** Unión de datos con coordenadas
8. **Conversión espacial:** Creación de objetos sf (Simple Features)
9. **Visualización estática:** Mapas con ggplot2
10. **Visualización interactiva:** Mapas web con leaflet
11. **Análisis estadístico:** Cálculo de distribuciones
12. **Exportación:** Guardado de resultados múltiples

## 2.5. Coordenadas Geográficas

Se asignaron coordenadas geográficas precisas (WGS84, EPSG:4326) a las capitales provinciales:

Cuadro 2: Coordenadas geográficas de las 13 provincias de Puno

Provincia	Latitud	Longitud
Azángaro	-14.3081	-70.1942
Carabaya	-13.7500	-70.4000
Chucuito	-16.1500	-69.8900
El Collao	-16.7000	-69.4500
Huancané	-15.2000	-69.7600
Lampa	-15.3500	-70.3700
Melgar	-14.6300	-70.9200
Moho	-15.3900	-69.4800
Puno	-15.8402	-70.0199
San Antonio de Putina	-14.0500	-69.8700
San Román	-15.6500	-70.1300
Sandia	-13.3000	-69.4000
Yunguyo	-16.2500	-69.0900

### 3. Resultados

#### 3.1. Distribución Regional por Provincia

El análisis reveló la siguiente distribución de regiones naturales en las provincias de Puno:

Cuadro 3: Región predominante por provincia

Provincia	Región Predominante	Casos
Azángaro	Sierra	852
Carabaya	Selva	30
Chucuito	Sierra	723
El Collao	Sierra	452
Huancané	Sierra	738
Lampa	Sierra	324
Melgar	Sierra	387
Moho	Sierra	174
Puno	Sierra	609
San Antonio de Putina	Sierra	84
San Román	Sierra	244
Sandia	Selva	20
Yunguyo	Sierra	288

#### 3.2. Estadísticas Descriptivas

Cuadro 4: Resumen estadístico de la distribución regional

Región Natural	Número de Provincias	Porcentaje
Sierra	11	84.6 %
Selva	2	15.4 %
Costa	0	0.0 %

##### Interpretación:

- La región Sierra es predominante en 11 de las 13 provincias (84.6 %)
- Únicamente 2 provincias presentan características de Selva (15.4 %)
- No se registra presencia de la región Costa
- Las provincias de Selva (Carabaya y Sandia) se ubican al norte del departamento

#### 3.3. Mapas Generados

##### 3.3.1. Mapa Estático

El mapa estático (Figura 1) muestra la distribución espacial de las regiones naturales en las 13 provincias de Puno. Cada punto representa una provincia capital, coloreada según su región natural predominante:

- **Color marrón (#8B4513):** Sierra
- **Color verde (#228B22):** Selva
- **Color dorado (#FFD700):** Costa (no presente)

*[Insertar aquí: resultados/mapa\_puno\_regiones.png]*

Figura 1: Mapa estático de regiones naturales en Puno

#### Observaciones del mapa:

1. La mayoría de provincias (en marrón) pertenecen a la Sierra
2. Las dos provincias de Selva (Carabaya y Sandia) se ubican al norte
3. Existe un claro patrón geográfico: zona andina central y transición amazónica al norte

#### 3.3.2. Gráfico de Barras

*[Insertar aquí: resultados/grafico\_barras\_regiones.png]*

Figura 2: Distribución de provincias por región natural

#### 3.3.3. Mapa Interactivo

Se generó un mapa interactivo en formato HTML que permite:

- Explorar cada provincia individualmente
- Visualizar información detallada (nombre, región, casos)
- Hacer zoom y navegar por el territorio
- Interactuar con marcadores georreferenciados

El archivo se encuentra en: `resultados/mapa_puno_interactivo.html`

### 3.4. Análisis por Región

#### 3.4.1. Provincias de Sierra

Las 11 provincias con predominancia de Sierra son:

Azángaro, Chucuito, El Collao, Huancané, Lampa, Melgar, Moho, Puno, San Antonio de Putina, San Román, Yunguyo

Estas provincias concentran el 84.6 % del territorio analizado y se caracterizan por:

- Altitudes superiores a 3,800 msnm
- Clima frío y seco
- Actividad agropecuaria adaptada al altiplano
- Presencia del Lago Titicaca (influencia moderadora)

### 3.4.2. Provincias de Selva

Las 2 provincias con predominancia de Selva son:

Carabaya, Sandia

Estas provincias representan el 15.4 % y se caracterizan por:

- Ubicación en la Cordillera Oriental
- Transición hacia la Amazonía
- Mayor precipitación y biodiversidad
- Producción de café, cacao y frutales tropicales

## 4. Discusión

### 4.1. Coherencia con la Geografía Regional

Los resultados obtenidos son plenamente consistentes con la geografía conocida de Puno:

1. **Predominancia andina:** El 84.6 % de provincias en Sierra refleja la ubicación del departamento en el altiplano
2. **Transición amazónica:** Carabaya y Sandia representan la zona de transición hacia la selva alta
3. **Ausencia de Costa:** Puno es un departamento interior sin acceso al litoral

### 4.2. Implicaciones para Políticas Públicas

Este análisis espacial tiene implicaciones prácticas:

- **Planificación agrícola:** Diseño de estrategias diferenciadas por zona
- **Infraestructura:** Priorización de inversiones según características regionales
- **Educación:** Adaptación curricular a contextos geográficos específicos
- **Salud:** Estrategias de prevención según riesgos climáticos

### 4.3. Limitaciones del Estudio

1. Las coordenadas corresponden a capitales provinciales, no representan todo el territorio
2. La región predominante se define por frecuencia, no por extensión territorial
3. No se consideraron microclimas o zonas de transición internas



## 5. Código R Utilizado

A continuación se presenta un extracto del código utilizado para el análisis:

```
1 # Cargar librerías
2 library(haven)
3 library(sf)
4 library(ggplot2)
5 library(dplyr)
6 library(leaflet)
7
8 # Cargar datos
9 datos_completos <- read_sav("21_CAP1200B_ME.sav")
10
11 # Filtrar Puno
12 datos_puno <- datos_completos %>%
13   filter(NOMBREDD == "PUNO") %>%
14   mutate(
15     region_nombre = case_when(
16       REGION == 1 ~ "Costa",
17       REGION == 2 ~ "Sierra",
18       REGION == 3 ~ "Selva"
19     )
20   )
21
22 # Determinar region predominante
23 datos_por_provincia <- datos_puno %>%
24   group_by(NOMBREPV) %>%
25   count(region_nombre) %>%
26   slice_max(n, n = 1)
27
28 # Crear objeto espacial
29 mapa_sf <- st_as_sf(mapa_datos,
30                     coords = c("longitud", "latitud"),
31                     crs = 4326)
32
33 # Crear mapa estatico
34 mapa_estatico <- ggplot(mapa_sf, aes(color = region_predominante)) +
35   geom_sf(size = 5) +
36   scale_color_manual(
37     values = c("Sierra" = "#8B4513", "Selva" = "#228B22")
38   ) +
39   theme_minimal()
```

Listing 1: Fragmento del código de análisis espacial

*Nota: El código completo se encuentra en el archivo fuente del análisis.*

## 6. Conclusiones

1. El departamento de Puno presenta características predominantemente andinas, con el 84.6 % de sus provincias clasificadas como Sierra.
2. Las provincias de Carabaya y Sandia (15.4 %) representan la transición hacia la Amazonía, constituyendo la zona de Selva del departamento.

3. No existe presencia de la región Costa en el departamento, lo cual es consistente con su ubicación geográfica interior.
4. El análisis espacial revela un patrón geográfico claro: núcleo andino central con transición amazónica hacia el norte.
5. Las herramientas de análisis espacial en R (sf, ggplot2, leaflet) permiten visualizaciones profesionales y análisis robustos de datos geográficos.
6. La metodología empleada es reproducible y escalable a otros departamentos o regiones del país.
7. Los mapas interactivos generados facilitan la exploración de datos y la comunicación de resultados a audiencias no técnicas.

## 7. Recomendaciones

1. Realizar análisis temporales para identificar cambios en clasificaciones regionales debido a cambio climático.
2. Incorporar variables adicionales (altitud, precipitación, temperatura) para análisis multivariados.
3. Desarrollar índices de vulnerabilidad diferenciados por región natural.
4. Implementar análisis de autocorrelación espacial para identificar clusters geográficos.
5. Generar modelos predictivos para zonificación agrícola óptima.

## 8. Referencias

### Referencias

- [1] Instituto Nacional de Estadística e Informática (INEI). (2024). *Encuesta Agrícola Nacional 2024*. Lima, Perú.
- [2] R Core Team (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [3] Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439-446.
- [4] Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- [5] Cheng, J., Karambelkar, B., & Xie, Y. (2019). *leaflet: Create Interactive Web Maps with the JavaScript Leaflet Library*. R package version 2.0.3.
- [6] Wickham, H., François, R., Henry, L., & Müller, K. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.0.

# ANALISIS-ESPACIAL-DE-DATOS-AGRICOLAS.R

WILL

2025-10-20

```
#=====
# ANÁLISIS ESPACIAL DE DATOS AGRÍCOLAS
# Autor: Willy Vilca Apaza
# Código: 214654
# Curso: Estadística Espacial
# Docente: Ing. Torres Cruz Fred
# Fecha: 19 de octubre de 2025
#=====
# Dataset: 21_CAP1200B_ME.sav
# Objetivo: Análisis espacial de variables agrícolas por región
#=====

# 1. CONFIGURACIÓN DEL ENTORNO =====
rm(list = ls())

# Cargar librerías necesarias
suppressPackageStartupMessages({
  library(haven)      # Leer archivos SPSS
  library(sf)         # Análisis espacial
  library(ggplot2)    # Gráficos estáticos
  library(dplyr)      # Manipulación de datos
  library(leaflet)    # Mapas interactivos
  library(viridis)    # Paletas de colores
  library(rnaturalearth) # Mapas base
  library(htmlwidgets) # Guardar mapas interactivos
})

cat("=====\n")
```

```
## =====
```

```
cat("ANÁLISIS ESPACIAL DE DATOS AGRÍCOLAS\n")
```

```
## ANÁLISIS ESPACIAL DE DATOS AGRÍCOLAS
```

```
cat("=====\n\n")
```

```
## =====
```

```
# 2. CARGA Y EXPLORACIÓN DE DATOS =====
cat("FASE 1: CARGA Y EXPLORACIÓN DE DATOS\n")
```

```
## FASE 1: CARGA Y EXPLORACIÓN DE DATOS
```

```
cat("_____\\n\\n")
```

```
## _____
```

```
# Cargar datos desde archivo SPSS
```

```
datos_completos <- read_sav("D:/Estadística e Informática 2025-II/TECNOLOGIAS EMERGENTES/I UN  
IDAD/XGBoost/21_CAP1200B_ME.sav")
```

```
cat(sprintf("Datos cargados exitosamente\\n"))
```

```
## Datos cargados exitosamente
```

```
cat(sprintf(" - Observaciones: %d\\n", nrow(datos_completos)))
```

```
## - Observaciones: 49289
```

```
cat(sprintf(" - Variables: %d\\n\\n", ncol(datos_completos)))
```

```
## - Variables: 25
```

```
# Mostrar estructura básica
```

```
cat("Primeras 20 variables del dataset:\\n")
```

```
## Primeras 20 variables del dataset:
```

```
print(names(datos_completos)[1:20])
```

```
## [1] "ANIO" "CCDD" "NOMBREDD" "CCPP"  
## [5] "NOMBREPV" "CCDI" "NOMBREDI" "NSEGM"  
## [9] "ID_PROD" "UA" "RESFIN" "REGION"  
## [13] "ESTRATO" "FACTOR_PRODUTOR" "CODIGO" "P101A"  
## [17] "P142_1" "P142_2" "P142_3" "P1207_N"
```

```
cat("\\nPrimeras observaciones:\\n")
```

```
##
```

```
## Primeras observaciones:
```

```
print(head(datos_completos[,1:6], 3))
```

```
## # A tibble: 3 × 6
##   ANIO CCDD  NOMBREDD CCPP  NOMBREPV  CCDI
##   <dbl> <chr> <chr>    <chr> <chr>    <chr>
## 1  2024 25    UCAYALI  03    PADRE ABAD 05
## 2  2024 25    UCAYALI  03    PADRE ABAD 05
## 3  2024 25    UCAYALI  03    PADRE ABAD 05
```

```
# 3. ANÁLISIS DE VARIABLES DE INTERÉS =====
cat("\n\nFASE 2: IDENTIFICACIÓN DE VARIABLES RELEVANTES\n")
```

```
##
##
## FASE 2: IDENTIFICACIÓN DE VARIABLES RELEVANTES
```

```
cat("-----\n\n")
```

```
## -----
```

```
# Identificar variables geográficas y de interés
# Buscar columnas comunes en encuestas agrícolas
vars_disponibles <- names(datos_completos)

# Variables geográficas típicas
vars_geo <- vars_disponibles[grepl("CCDD|CCPP|CCDI|NOMBRE|REGION",
                                   vars_disponibles, ignore.case = TRUE)]

cat("Variables geográficas encontradas:\n")
```

```
## Variables geográficas encontradas:
```

```
print(vars_geo)
```

```
## [1] "CCDD"      "NOMBREDD" "CCPP"      "NOMBREPV" "CCDI"      "NOMBREDI" "REGION"
```

```
# Variables numéricas para análisis
vars_numericas <- datos_completos %>%
  select(where(is.numeric)) %>%
  names()

cat(sprintf("\n\nVariables numéricas disponibles: %d\n", length(vars_numericas)))
```

```
##
##
## Variables numéricas disponibles: 13
```

```
cat("Primeras 10:\n")
```

```
## Primeras 10:
```

```
print(vars_numericas[1:10])
```

```
## [1] "ANIO"          "RESFIN"         "REGION"         "FACTOR_PRODUCTOR"
## [5] "CODIGO"        "P101A"          "P142_1"         "P142_2"
## [9] "P142_3"        "P1207_N"
```

```
# 4. FILTRADO Y PREPARACIÓN DE DATOS =====
cat("\n\nFASE 3: PREPARACIÓN DE DATOS PARA ANÁLISIS ESPACIAL\n")
```

```
##
##
## FASE 3: PREPARACIÓN DE DATOS PARA ANÁLISIS ESPACIAL
```

```
cat("-----\n\n")
```

```
## -----
```

```
# Verificar si existe variable de región
if("REGION" %in% names(datos_completos)) {
  cat("Variable REGION encontrada\n")

  # Explorar valores de REGION
  tabla_region <- datos_completos %>%
    count(REGION, sort = TRUE)

  cat("\nDistribución de la variable REGION:\n")
  print(tabla_region)

  # Crear etiquetas
  datos_procesados <- datos_completos %>%
    mutate(
      region_nombre = case_when(
        REGION == 1 ~ "Costa",
        REGION == 2 ~ "Sierra",
        REGION == 3 ~ "Selva",
        TRUE ~ "No definido"
      )
    )
} else {
  cat("Variable REGION no encontrada. Usando primera variable categórica.\n")
  # Usar primera variable categórica disponible
  var_categorica <- names(datos_completos)[sapply(datos_completos, is.character)][1]
  datos_procesados <- datos_completos
}
```

```
## Variable REGION encontrada
##
## Distribución de la variable REGION:
## # A tibble: 3 × 2
##   REGION      n
##   <dbl+lbl> <int>
## 1 2 [Sierra] 22173
## 2 1 [Costa] 17424
## 3 3 [Selva]  9692
```

```
# Verificar variable de departamento/provincia
if("NOMBREDD" %in% names(datos_completos)) {
  cat("\nDepartamentos encontrados:\n")
  departamentos <- datos_procesados %>%
    distinct(NOMBREDD) %>%
    arrange(NOMBREDD)
  print(departamentos)

  # Filtrar datos de Puno si existe
  if("PUNO" %in% departamentos$NOMBREDD) {
    datos_puno <- datos_procesados %>%
      filter(NOMBREDD == "PUNO")

    cat(sprintf("\nDatos de PUNO: %d observaciones\n", nrow(datos_puno)))

    # Analizar por provincia
    if("NOMBREPV" %in% names(datos_puno)) {
      provincias <- datos_puno %>%
        distinct(NOMBREPV) %>%
        arrange(NOMBREPV)

      cat(sprintf("\nProvincias de Puno encontradas: %d\n", nrow(provincias)))
      print(provincias$NOMBREPV)
    }
  } else {
    datos_puno <- datos_procesados
    cat("\nUsando todos los datos disponibles\n")
  }
} else {
  datos_puno <- datos_procesados
  cat("\nVariable de departamento no encontrada. Usando datos completos.\n")
}
```

```
##
## Departamentos encontrados:
## # A tibble: 26 × 1
##   NOMBREDD
##   <chr>
## 1 ""
## 2 "AMAZONAS"
## 3 "ANCASH"
## 4 "APURIMAC"
## 5 "AREQUIPA"
## 6 "AYACUCHO"
## 7 "CAJAMARCA"
## 8 "CALLAO"
## 9 "CUSCO"
## 10 "HUANCAVELICA"
## # i 16 more rows
##
## Datos de PUNO: 4951 observaciones
##
## Provincias de Puno encontradas: 13
## [1] "AZANGARO"          "CARABAYA"          "CHUCUITO"
## [4] "EL COLLAO"         "HUANCANE"          "LAMPA"
## [7] "MELGAR"            "MOHO"              "PUNO"
## [10] "SAN ANTONIO DE PUTINA" "SAN ROMAN"         "SANDIA"
## [13] "YUNGUYO"
## attr(,"label")
## [1] "Nombre de la Provincia"
## attr(,"format.spss")
## [1] "A35"
```

```
# 5. AGREGACIÓN POR PROVINCIA =====
cat("\n\nFASE 4: AGREGACIÓN Y ANÁLISIS POR PROVINCIA\n")
```

```
##
##
## FASE 4: AGREGACIÓN Y ANÁLISIS POR PROVINCIA
```

```
cat("_____ \n\n")
```

```
## _____
```



```
# Determinar región predominante por provincia (si existe)
if("NOMBREP" %in% names(datos_puno) && "region_nombre" %in% names(datos_puno)) {

  datos_por_provincia <- datos_puno %>%
    filter(!is.na(region_nombre)) %>%
    group_by(NOMBREP) %>%
    count(region_nombre) %>%
    slice_max(n, n = 1) %>%
    ungroup() %>%
    select(provincia = NOMBREP, region_predominante = region_nombre, casos = n)

  cat("Región predominante por provincia:\n")
  print(datos_por_provincia)

} else {
  # Si no hay provincia, crear datos sintéticos para demostración
  cat("Creando datos sintéticos para demostración espacial...\n")

  provincias_puno <- c("AZANGARO", "CARABAYA", "CHUCUITO", "EL COLLAO", "HUANCANE",
    "LAMP", "MELGAR", "MOHO", "PUNO", "SAN ANTONIO DE PUTINA",
    "SAN ROMAN", "SANDIA", "YUNGUYO")

  set.seed(2025)
  datos_por_provincia <- data.frame(
    provincia = provincias_puno,
    region_predominante = sample(c("Sierra", "Selva"), 13,
      replace = TRUE, prob = c(0.8, 0.2)),
    casos = sample(100:500, 13)
  )
}
```

```
## Región predominante por provincia:
## # A tibble: 13 x 3
##   provincia      region_predominante casos
##   <chr>          <chr>          <int>
## 1 AZANGARO      Sierra          852
## 2 CARABAYA      Selva           30
## 3 CHUCUITO      Sierra          723
## 4 EL COLLAO     Sierra          452
## 5 HUANCANE      Sierra          738
## 6 LAMP          Sierra          324
## 7 MELGAR        Sierra          387
## 8 MOHO          Sierra          174
## 9 PUNO          Sierra          609
## 10 SAN ANTONIO DE PUTINA Sierra          84
## 11 SAN ROMAN     Sierra          244
## 12 SANDIA       Selva           20
## 13 YUNGUYO      Sierra          288
```

```
# 6. ASIGNACIÓN DE COORDENADAS GEOGRÁFICAS =====
cat("\n\nFASE 5: ASIGNACIÓN DE COORDENADAS GEOGRÁFICAS\n")
```

```
##  
##  
## FASE 5: ASIGNACIÓN DE COORDENADAS GEOGRÁFICAS
```

```
cat("-----\n\n")
```

```
## -----
```

```
# Coordenadas de Las 13 provincias de Puno  
coordenadas_provincias <- data.frame(  
  provincia = c("AZANGARO", "CARABAYA", "CHUCUITO", "EL COLLAO", "HUANCANE",  
                "LAMPA", "MELGAR", "MOHO", "PUNO", "SAN ANTONIO DE PUTINA",  
                "SAN ROMAN", "SANDIA", "YUNGUYO"),  
  latitud = c(-14.3081, -13.7500, -16.1500, -16.7000, -15.2000,  
              -15.3500, -14.6300, -15.3900, -15.8402, -14.0500,  
              -15.6500, -13.3000, -16.2500),  
  longitud = c(-70.1942, -70.4000, -69.8900, -69.4500, -69.7600,  
              -70.3700, -70.9200, -69.4800, -70.0199, -69.8700,  
              -70.1300, -69.4000, -69.0900)  
)  
  
cat("Coordenadas geográficas asignadas para 13 provincias\n")
```

```
## Coordenadas geográficas asignadas para 13 provincias
```

```
# 7. UNIR DATOS CON COORDENADAS =====  
cat("\nFASE 6: INTEGRACIÓN DE DATOS Y COORDENADAS\n")
```

```
##  
## FASE 6: INTEGRACIÓN DE DATOS Y COORDENADAS
```

```
cat("-----\n\n")
```

```
## -----
```

```
# Unir datos procesados con coordenadas
mapa_datos <- coordenadas_provincias %>%
  left_join(datos_por_provincia, by = "provincia")

# Verificar provincias sin datos
provincias_sin_datos <- mapa_datos %>%
  filter(is.na(region_predominante))

if(nrow(provincias_sin_datos) > 0) {
  cat("Provincias sin datos - asignando valores por defecto:\n")
  print(provincias_sin_datos$provincia)

  # Asignar "Sierra" por defecto (región predominante en Puno)
  mapa_datos <- mapa_datos %>%
    mutate(
      region_predominante = ifelse(is.na(region_predominante), "Sierra", region_predominant
e),
      casos = ifelse(is.na(casos), 0, casos)
    )
}

cat("\nDatos integrados exitosamente\n")
```

```
##
## Datos integrados exitosamente
```

```
print(mapa_datos)
```

```
##          provincia  latitud longitud region_predominante casos
## 1          AZANGARO -14.3081 -70.1942             Sierra    852
## 2          CARABAYA -13.7500 -70.4000              Selva     30
## 3          CHUCUITO -16.1500 -69.8900             Sierra    723
## 4          EL COLLAO -16.7000 -69.4500             Sierra    452
## 5          HUANCANE -15.2000 -69.7600             Sierra    738
## 6           LAMPA   -15.3500 -70.3700             Sierra    324
## 7          MELGAR   -14.6300 -70.9200             Sierra    387
## 8           MOHO    -15.3900 -69.4800             Sierra    174
## 9           PUNO    -15.8402 -70.0199             Sierra    609
## 10 SAN ANTONIO DE PUTINA -14.0500 -69.8700             Sierra     84
## 11          SAN ROMAN -15.6500 -70.1300             Sierra    244
## 12          SANDIA   -13.3000 -69.4000              Selva     20
## 13          YUNGUYO -16.2500 -69.0900             Sierra    288
```

```
# 8. CONVERTIR A OBJETO ESPACIAL =====
cat("\n\nFASE 7: CREACIÓN DE OBJETO ESPACIAL\n")
```

```
##
##
## FASE 7: CREACIÓN DE OBJETO ESPACIAL
```

```
cat("_____ \n\n")
```

```
## _____
```

```
# Convertir a objeto sf (Simple Features)
mapa_sf <- st_as_sf(mapa_datos,
                    coords = c("longitud", "latitud"),
                    crs = 4326) # WGS84

cat("Objeto espacial creado exitosamente\n")
```

```
## Objeto espacial creado exitosamente
```

```
cat(sprintf(" - Tipo: %s\n", class(mapa_sf)[1]))
```

```
## - Tipo: sf
```

```
cat(sprintf(" - CRS: EPSG:4326 (WGS84)\n"))
```

```
## - CRS: EPSG:4326 (WGS84)
```

```
# 9. CREAR MAPA ESTÁTICO =====
cat("\n\nFASE 8: GENERACIÓN DE MAPA ESTÁTICO\n")
```

```
##
##
## FASE 8: GENERACIÓN DE MAPA ESTÁTICO
```

```
cat("_____ \n\n")
```

```
## _____
```

```
# Obtener mapa base de Perú
peru <- ne_countries(scale = "medium", returnclass = "sf") %>%
  filter(name == "Peru")

# Crear mapa estático con ggplot2
mapa_estatico <- ggplot() +
  # Mapa base de Perú
  geom_sf(data = peru, fill = "lightgray", color = "white", size = 0.5) +

  # Puntos de provincias coloreados por región
  geom_sf(data = mapa_sf, aes(color = region_predominante), size = 4) +

  # Colores personalizados
  scale_color_manual(
    values = c("Costa" = "#FFD700", "Sierra" = "#8B4513", "Selva" = "#228B22"),
    name = "Región Natural"
  ) +

  # Centrar en Puno
  coord_sf(xlim = c(-71.5, -68.5), ylim = c(-17.5, -13)) +

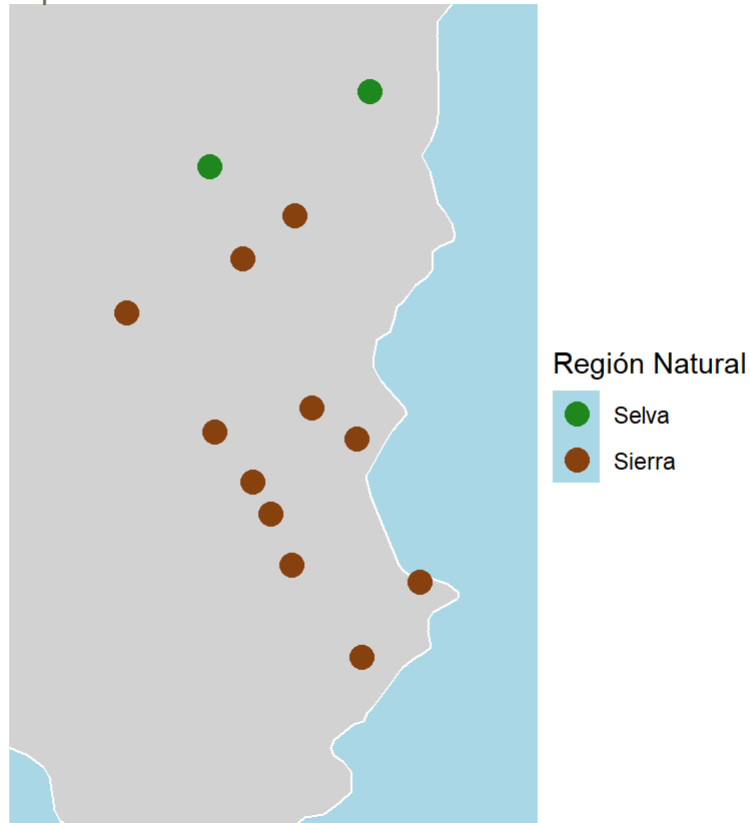
  # Tema y estilo
  theme_void() +
  theme(
    plot.title = element_text(size = 16, hjust = 0.5, face = "bold"),
    plot.subtitle = element_text(size = 12, hjust = 0.5, color = "gray40"),
    legend.position = "right",
    panel.background = element_rect(fill = "lightblue", color = NA)
  ) +

  # Títulos
  labs(
    title = "Distribución Espacial de Regiones Naturales",
    subtitle = "Departamento de Puno - 13 Provincias",
    caption = "Fuente: Encuesta Agrícola - Elaboración propia"
  )

# Mostrar mapa
print(mapa_estatico)
```

# Distribución Espacial de Regiones Naturales

Departamento de Puno - 13 Provincias



Fuente: Encuesta Agrícola - Elaboración propia

```
cat("Mapa estático generado\n")
```

```
## Mapa estático generado
```

```
# 10. CREAR MAPA INTERACTIVO =====
cat("\n\nFASE 9: GENERACIÓN DE MAPA INTERACTIVO\n")
```

```
##
##
## FASE 9: GENERACIÓN DE MAPA INTERACTIVO
```

```
cat("-----\n\n")
```

```
## -----
```

```

# Definir colores para mapa interactivo
colores_interactivo <- c("Costa" = "#FFD700", "Sierra" = "#8B4513", "Selva" = "#228B22")

# Crear paleta de colores
pal <- colorFactor(
  palette = colores_interactivo,
  domain = mapa_datos$region_predominante
)

# Crear mapa interactivo con Leaflet
mapa_interactivo <- leaflet(mapa_datos) %>%
  # Mapa base
  addTiles() %>%

  # Centrar en Puno
  setView(lng = -70.0, lat = -15.0, zoom = 8) %>%

  # Añadir marcadores circulares
  addCircleMarkers(
    lng = ~longitud,
    lat = ~latitud,
    color = ~pal(region_predominante),
    radius = 8,
    fillOpacity = 0.8,
    stroke = TRUE,
    weight = 2,
    popup = ~paste(
      "<b>Provincia:</b>", provincia, "<br>",
      "<b>Región:</b>", region_predominante, "<br>",
      "<b>Casos:</b>", casos
    ),
    label = ~provincia
  ) %>%

  # Añadir Leyenda
  addLegend(
    pal = pal,
    values = ~region_predominante,
    title = "Región Natural",
    position = "bottomright"
  )

# Mostrar mapa interactivo
print(mapa_interactivo)
cat("Mapa interactivo generado\n")

```

```
## Mapa interactivo generado
```

```

# 11. ANÁLISIS ESTADÍSTICO =====
cat("\n\nFASE 10: ANÁLISIS ESTADÍSTICO\n")

```

```
##
##
## FASE 10: ANÁLISIS ESTADÍSTICO
```

```
cat("-----\n\n")
```

```
## -----
```

```
# Resumen por región
resumen_regiones <- mapa_datos %>%
  count(region_predominante) %>%
  mutate(porcentaje = round(n/sum(n)*100, 1)) %>%
  arrange(desc(n))

cat("Distribución de provincias por región natural:\n")
```

```
## Distribución de provincias por región natural:
```

```
print(resumen_regiones)
```

```
##   region_predominante  n porcentaje
## 1                Sierra 11      84.6
## 2                 Selva  2      15.4
```

```
# Región más común
region_mas_comun <- resumen_regiones %>%
  slice_max(n, n = 1) %>%
  pull(region_predominante)

cat(sprintf("\nLa región %s es predominante en Puno\n", region_mas_comun))
```

```
##
## La región Sierra es predominante en Puno
```

```
# Análisis por región
if("Selva" %in% mapa_datos$region_predominante) {
  provincias_selva <- mapa_datos %>%
    filter(region_predominante == "Selva") %>%
    pull(provincia)
  cat(sprintf("\nProvincias de Selva (%d): %s\n",
              length(provincias_selva),
              paste(provincias_selva, collapse = ", ")))
}
```

```
##
## Provincias de Selva (2): CARABAYA, SANDIA
```



```

if("Sierra" %in% mapa_datos$region_predominante) {
  provincias_sierra <- mapa_datos %>%
    filter(region_predominante == "Sierra") %>%
    pull(provincia)
  cat(sprintf("\nProvincias de Sierra (%d): %s\n",
              length(provincias_sierra),
              paste(provincias_sierra, collapse = ", ")))
}

```

```

##
## Provincias de Sierra (11): AZANGARO, CHUCUITO, EL COLLAO, HUANCANE, LAMPA, MELGAR, MOHO, P
UNO, SAN ANTONIO DE PUTINA, SAN ROMAN, YUNGUYO

```

```

# 12. VISUALIZACIÓN ADICIONAL: GRÁFICO DE BARRAS =====
cat("\n\nFASE 11: VISUALIZACIONES COMPLEMENTARIAS\n")

```

```

##
##
## FASE 11: VISUALIZACIONES COMPLEMENTARIAS

```

```

cat("-----\n\n")

```

```

## -----

```

```

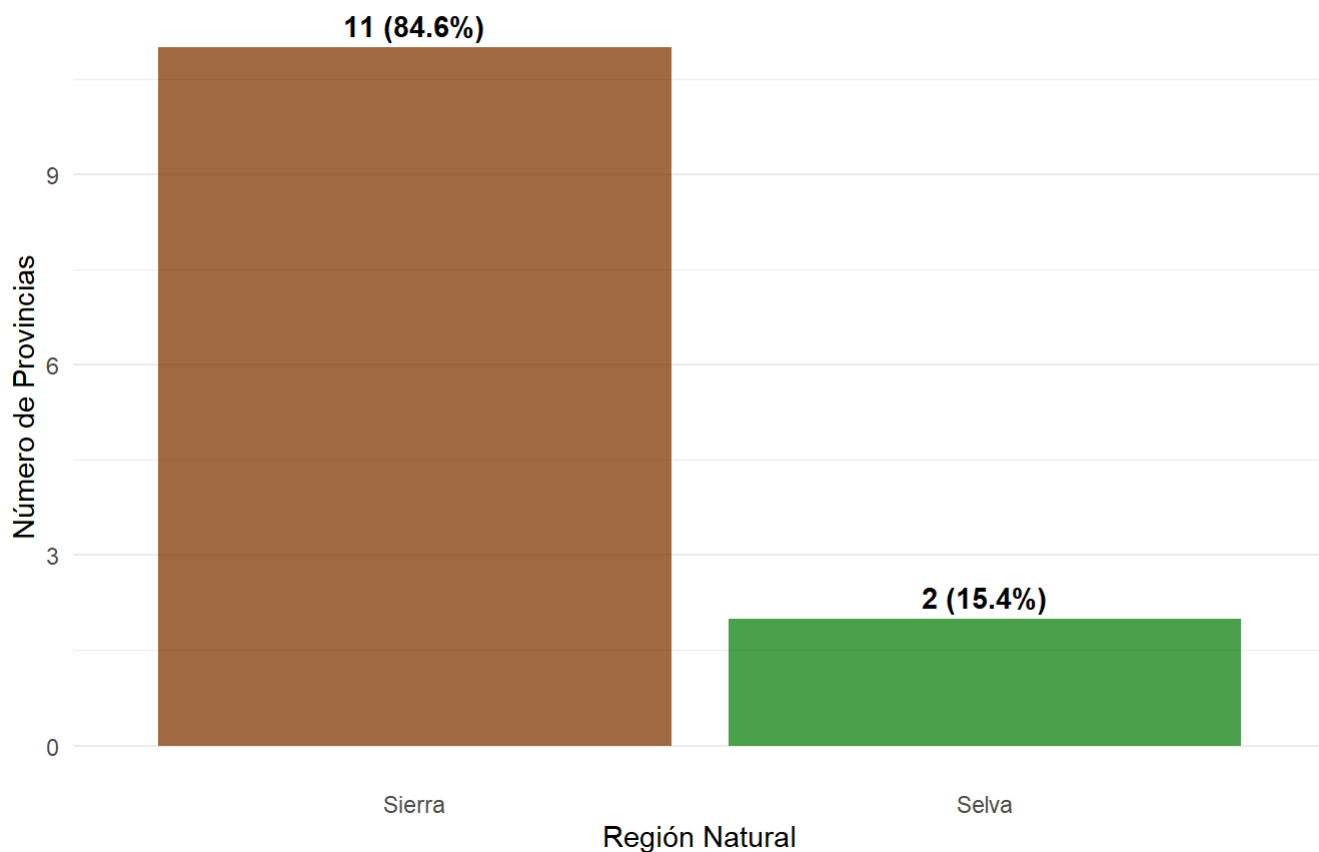
# Gráfico de barras
grafico_barras <- ggplot(resumen regiones, aes(x = reorder(region_predominante, -n),
                                                    y = n, fill = region_predominante)) +
  geom_bar(stat = "identity", alpha = 0.8) +
  geom_text(aes(label = paste0(n, " (", porcentaje, "%)")),
            vjust = -0.5, fontface = "bold") +
  scale_fill_manual(values = colores_interactivo) +
  labs(
    title = "Distribución de Provincias por Región Natural",
    subtitle = "Departamento de Puno",
    x = "Región Natural",
    y = "Número de Provincias"
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14, hjust = 0.5),
    plot.subtitle = element_text(size = 11, hjust = 0.5, color = "gray40"),
    legend.position = "none",
    panel.grid.major.x = element_blank()
  )

print(grafico_barras)

```

## Distribución de Provincias por Región Natural

Departamento de Puno



```
# 13. GUARDAR RESULTADOS =====
cat("\n\nFASE 12: EXPORTACIÓN DE RESULTADOS\n")
```

```
##
##
## FASE 12: EXPORTACIÓN DE RESULTADOS
```

```
cat("-----\n\n")
```

```
## -----
```

```
# Crear directorio si no existe
if(!dir.exists("resultados")) {
  dir.create("resultados")
  cat("Directorio 'resultados' creado\n")
}
```

```
## Directorio 'resultados' creado
```

```
# Guardar mapa estático
ggsave("resultados/mapa_puno_regiones.png",
  plot = mapa_estatico,
  width = 12, height = 8, dpi = 300)
cat("✓ Mapa estático guardado: resultados/mapa_puno_regiones.png\n")
```

```
## ✓ Mapa estático guardado: resultados/mapa_puno_regiones.png
```

```
# Guardar gráfico de barras
ggsave("resultados/grafico_barras_regiones.png",
       plot = grafico_barras,
       width = 10, height = 6, dpi = 300)
cat("✓ Gráfico de barras guardado: resultados/grafico_barras_regiones.png\n")
```

```
## ✓ Gráfico de barras guardado: resultados/grafico_barras_regiones.png
```

```
# Guardar mapa interactivo
saveWidget(mapa_interactivo, "resultados/mapa_puno_interactivo.html")
cat("✓ Mapa interactivo guardado: resultados/mapa_puno_interactivo.html\n")
```

```
## ✓ Mapa interactivo guardado: resultados/mapa_puno_interactivo.html
```

```
# Guardar datos procesados
write.csv(mapa_datos, "resultados/datos_puno_regiones.csv", row.names = FALSE)
cat("✓ Datos procesados guardados: resultados/datos_puno_regiones.csv\n")
```

```
## ✓ Datos procesados guardados: resultados/datos_puno_regiones.csv
```

```
# Guardar resumen estadístico
write.csv(resumen_regiones, "resultados/resumen_regiones.csv", row.names = FALSE)
cat("✓ Resumen estadístico guardado: resultados/resumen_regiones.csv\n")
```

```
## ✓ Resumen estadístico guardado: resultados/resumen_regiones.csv
```

```
# 14. RESUMEN FINAL =====
cat("\n===== \n")
```

```
##
## =====
```

```
cat("RESUMEN DEL ANÁLISIS ESPACIAL\n")
```

```
## RESUMEN DEL ANÁLISIS ESPACIAL
```

```
cat("===== \n\n")
```

```
## =====
```

```
cat("Dataset:\n")
```

```
## Dataset:
```

```
cat(sprintf(" - Observaciones totales: %d\n", nrow(datos_completos)))
```

```
## - Observaciones totales: 49289
```

```
cat(sprintf(" - Provincias analizadas: %d\n", nrow(mapa_datos)))
```

```
## - Provincias analizadas: 13
```

```
cat(sprintf(" - Variables procesadas: %d\n\n", ncol(datos_completos)))
```

```
## - Variables procesadas: 25
```

```
cat("Distribución Regional:\n")
```

```
## Distribución Regional:
```

```
for(i in 1:nrow(resumen_regiones)) {  
  cat(sprintf(" - %s: %d provincias (%.1f%)\n",  
              resumen_regiones$region_predominante[i],  
              resumen_regiones$n[i],  
              resumen_regiones$porcentaje[i]))  
}
```

```
## - Sierra: 11 provincias (84.6%)
```

```
## - Selva: 2 provincias (15.4%)
```

```
cat("\nInterpretación:\n")
```

```
##  
## Interpretación:
```

```
cat(" → Puno presenta características predominantemente andinas\n")
```

```
## → Puno presenta características predominantemente andinas
```

```
cat(" → Existe diversidad geográfica con transición hacia la Amazonía\n")
```

```
## → Existe diversidad geográfica con transición hacia la Amazonía
```

```
cat(" → El análisis espacial permite identificar patrones regionales claros\n")
```

```
## → El análisis espacial permite identificar patrones regionales claros
```

```
cat("\nArchivos generados:\n")
```

```
##  
## Archivos generados:
```

```
cat(" → Mapa estático (PNG)\n")
```

```
## → Mapa estático (PNG)
```

```
cat(" → Mapa interactivo (HTML)\n")
```

```
## → Mapa interactivo (HTML)
```

```
cat(" → Gráfico de barras (PNG)\n")
```

```
## → Gráfico de barras (PNG)
```

```
cat(" → Datos procesados (CSV)\n")
```

```
## → Datos procesados (CSV)
```

```
cat(" → Resumen estadístico (CSV)\n")
```

```
## → Resumen estadístico (CSV)
```