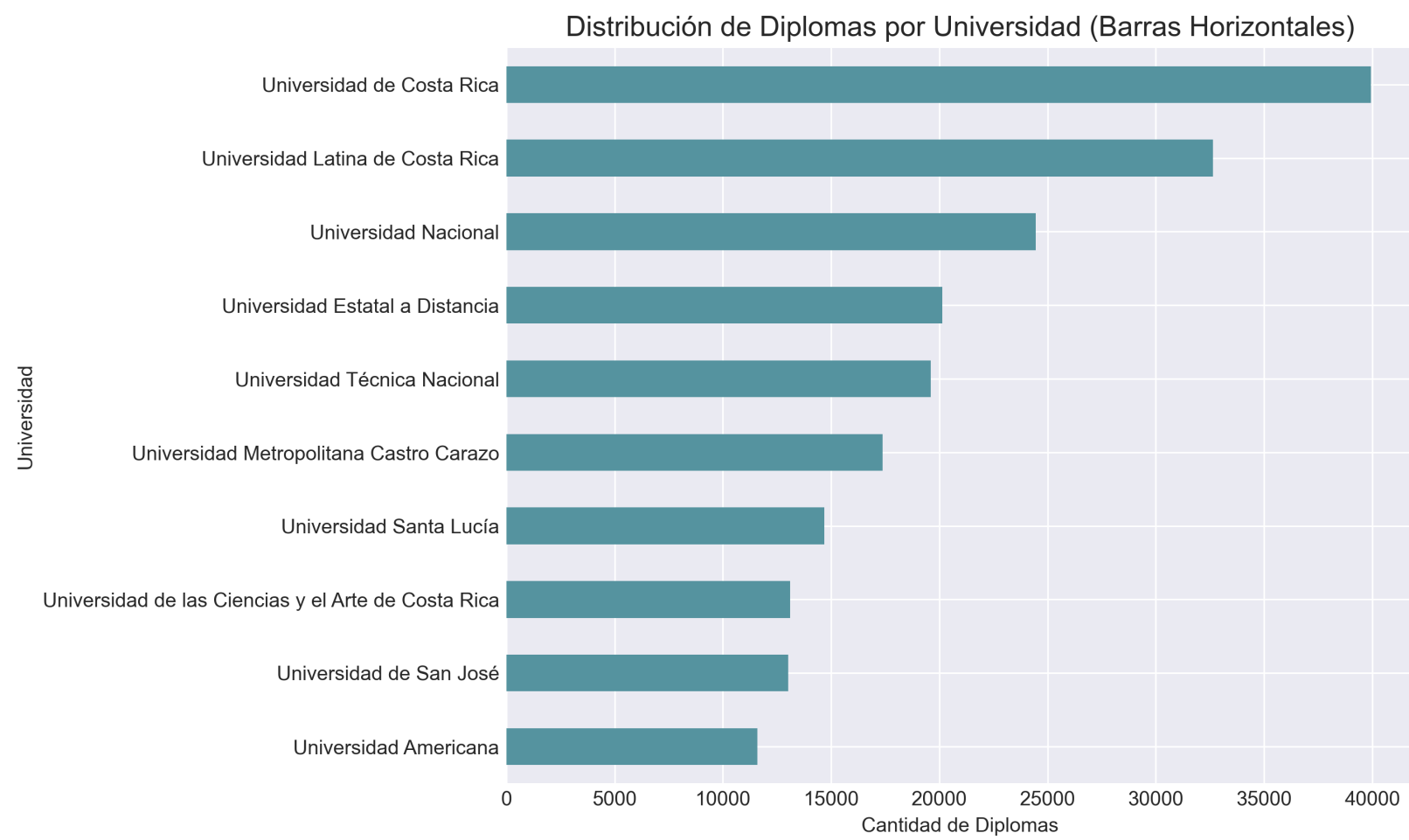


Guía Visual de Gráficos Estadísticos en Python

Trabajo Corto #2: Gráficos para Estadística Descriptiva
Colegio Universitario de Cartago | Kendal Solano y Roberto Coto
Big Data - III Cuatrimestre 2025 | Prof. David Martínez Salazar

1. Gráfico de Barras Horizontales

Definición: Barras horizontales donde la longitud representa el valor de cada categoría.
Cuando usar: Series cualitativas o geográficas con nombres largos.
Ejemplo: Top 10 universidades por diplomas emitidos (2014-2020)

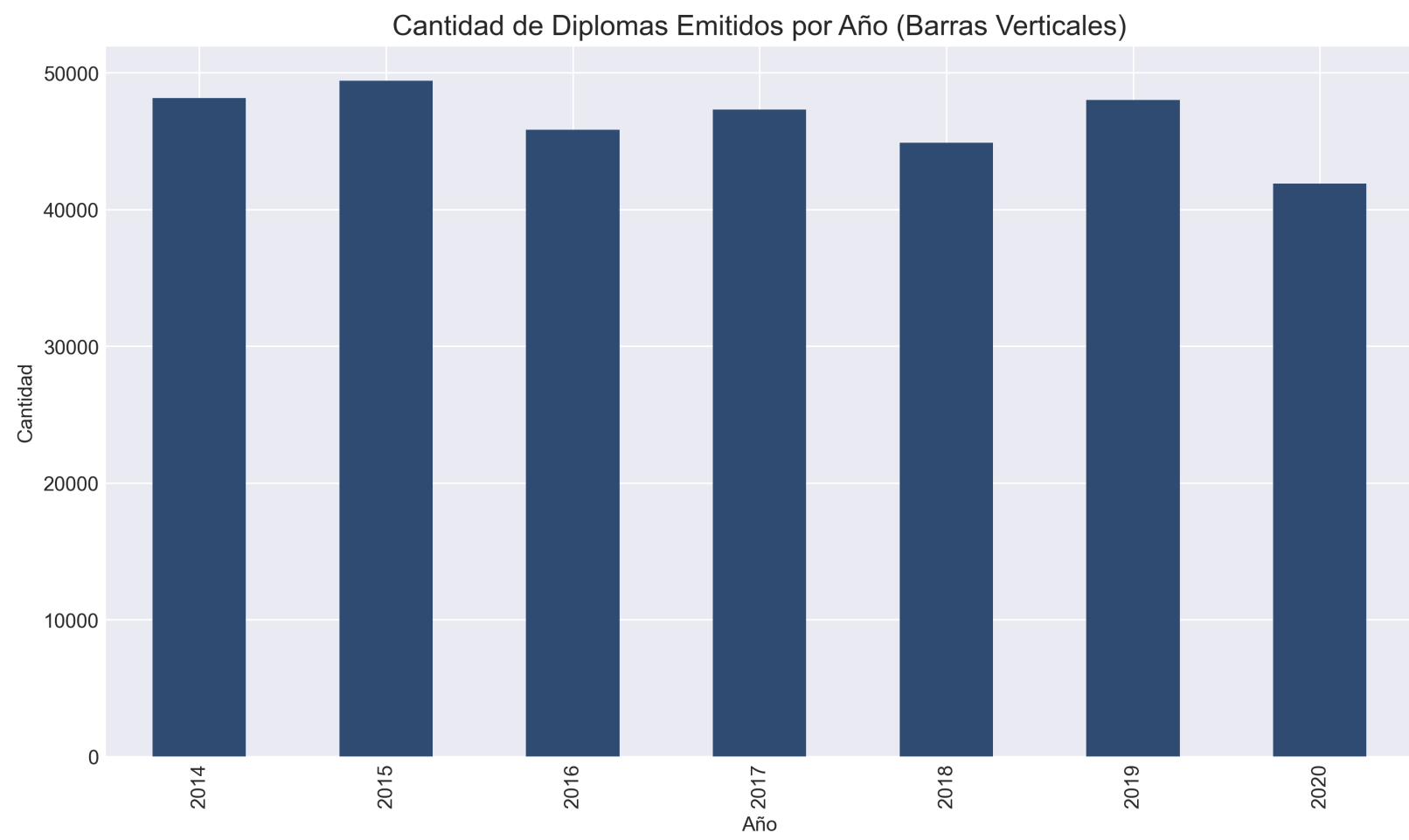


Universidad de Costa Rica lidera con >40,000 diplomas

```
df.groupby('Universidad').sum()['Cantidad']  
df.sort_values('Cantidad', ascending=False).head(10)  
df.groupby('Universidad').mean()['Cantidad']
```

2. Gráfico de Barras Verticales

Definición: Barras que crecen verticalmente desde el eje horizontal.
Cuando usar: Comparar categorías discretas, especialmente datos temporales.
Ejemplo: Diplomas emitidos por año (2014-2020)

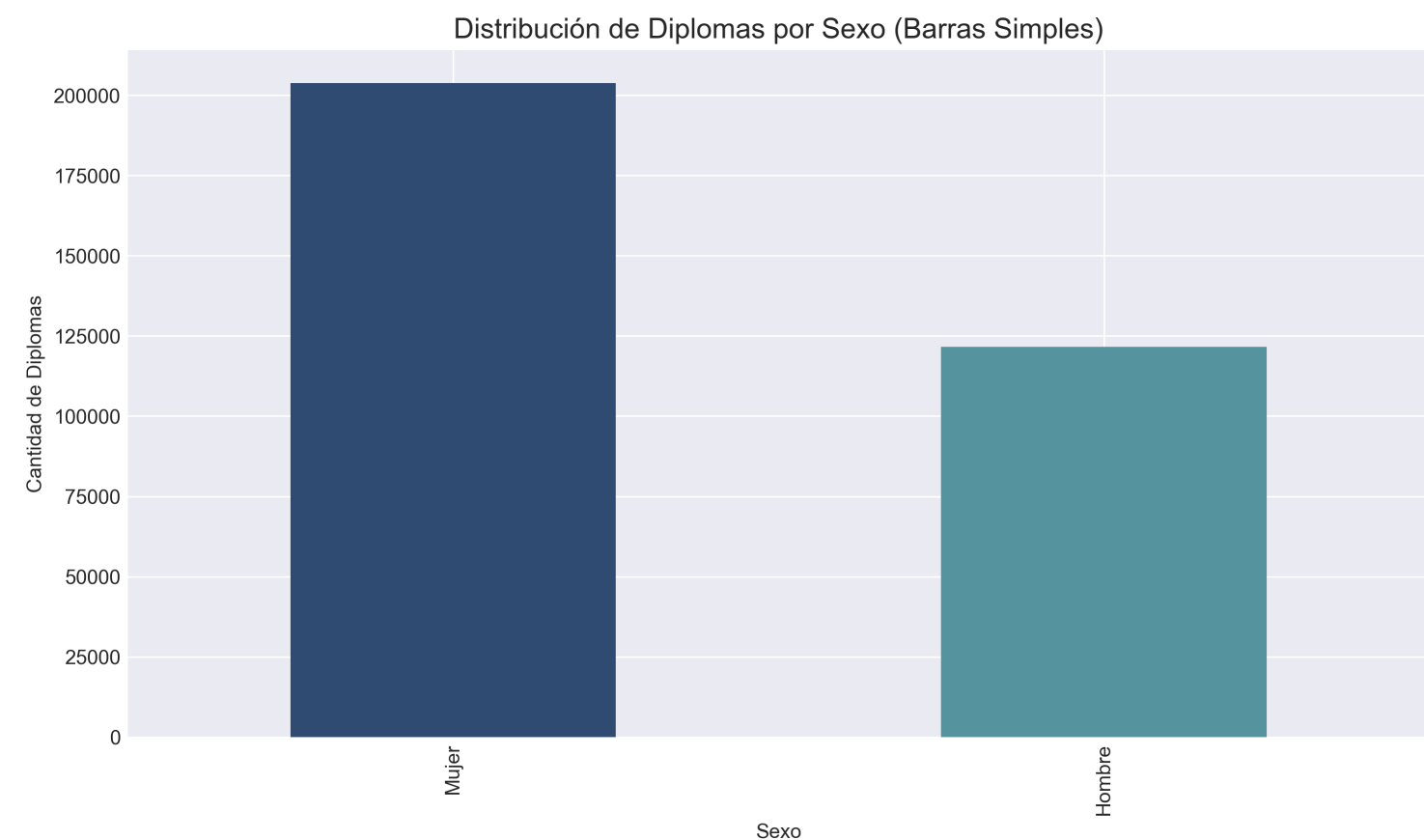


Pico en 2015 (49,500), caída en 2020 (41,900)

```
df.groupby('Año').sum()['Cantidad']  
df.groupby('Año').mean()['Cantidad']  
df.groupby('Año').std()['Cantidad']
```

3. Gráfico de Barras Simples

Definición: Representación básica de frecuencias para una variable cualitativa.
Cuando usar: Comparar categorías mutuamente excluyentes.
Ejemplo: Distribución de diplomas por sexo

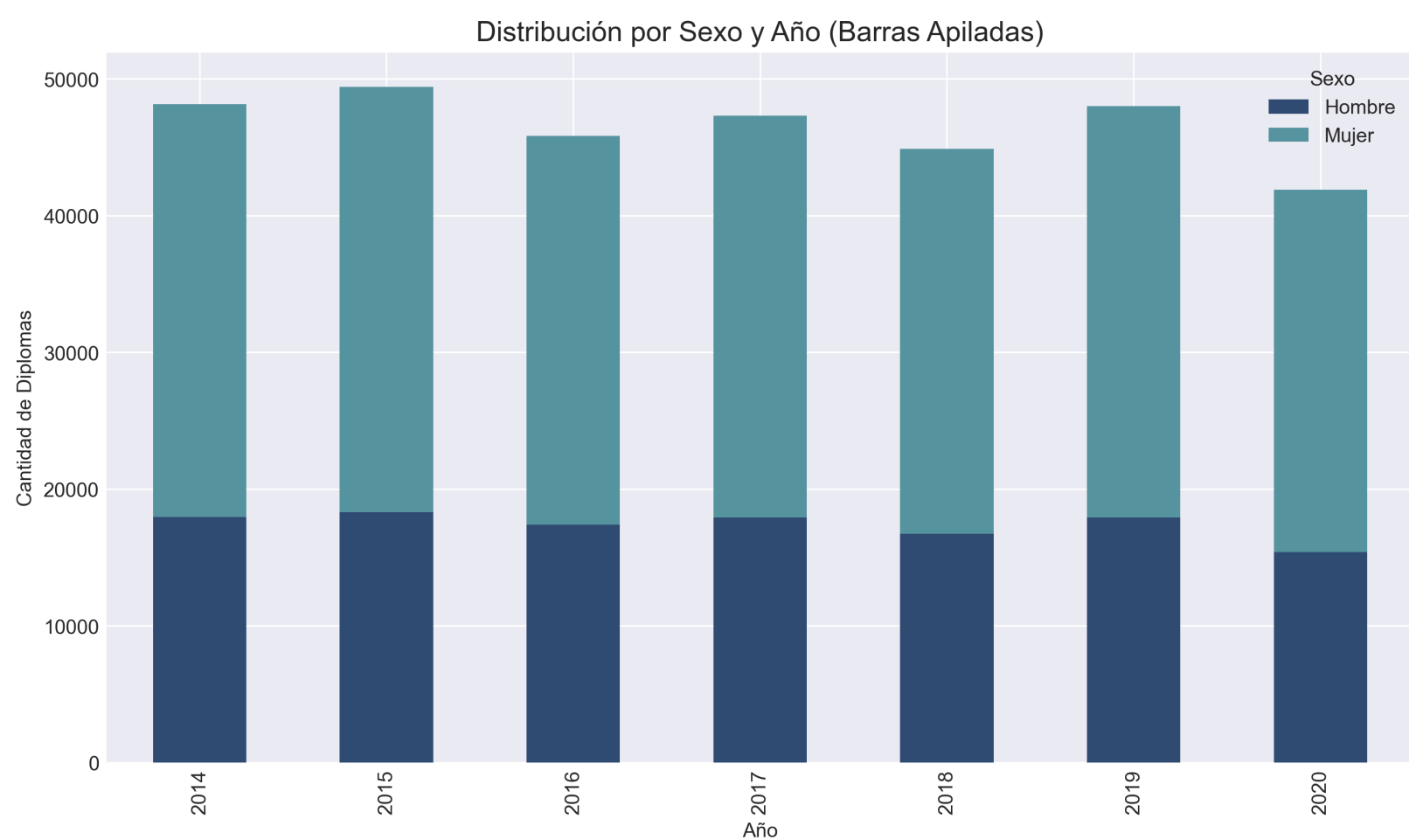


Mujeres: 62% (204,000) | Hombres: 38% (121,000)

```
df.groupby('Sexo').sum()['Cantidad']  
df.groupby('Sexo').mean()['Cantidad']  
df.groupby('Sexo').std()['Cantidad']
```

4. Gráfico de Barras Apiladas

Definición: Barras divididas en segmentos apilados que muestran composición interna.
Cuando usar: Comparar totales y ver composición simultáneamente.
Ejemplo: Diplomas por año diferenciando sexo (valores absolutos)

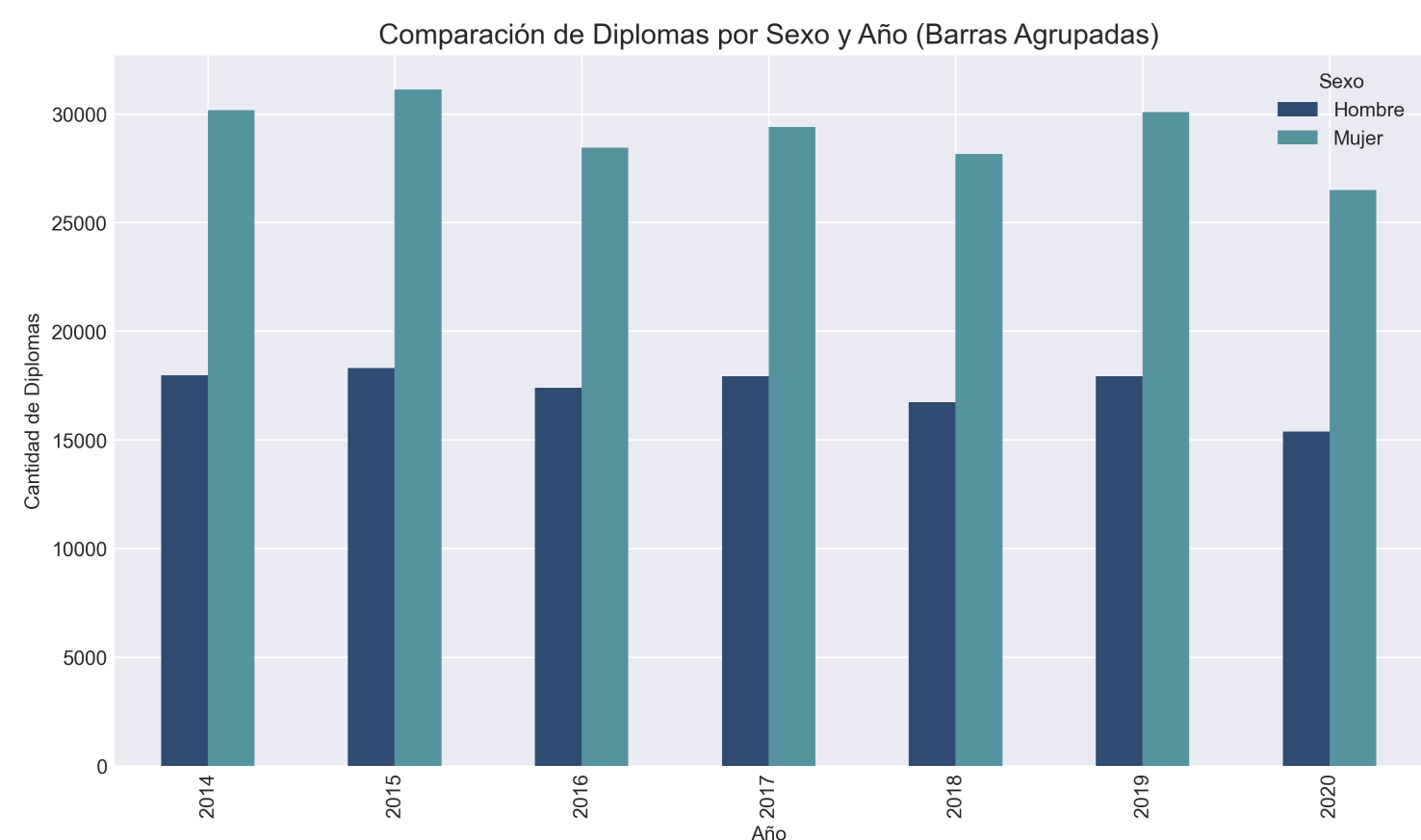


Dominancia femenina consistente en todos los años

```
df.groupby('Año').sum()['Cantidad']  
df.groupby('Año').sum()['Cantidad']  
df.groupby('Año').sum()['Cantidad']
```

5. Gráfico de Barras Agrupadas

Definición: Barras lado a lado para comparación directa entre subcategorías.
Cuando usar: Comparar magnitudes absolutas entre grupos.
Ejemplo: Comparación directa hombres vs mujeres por año

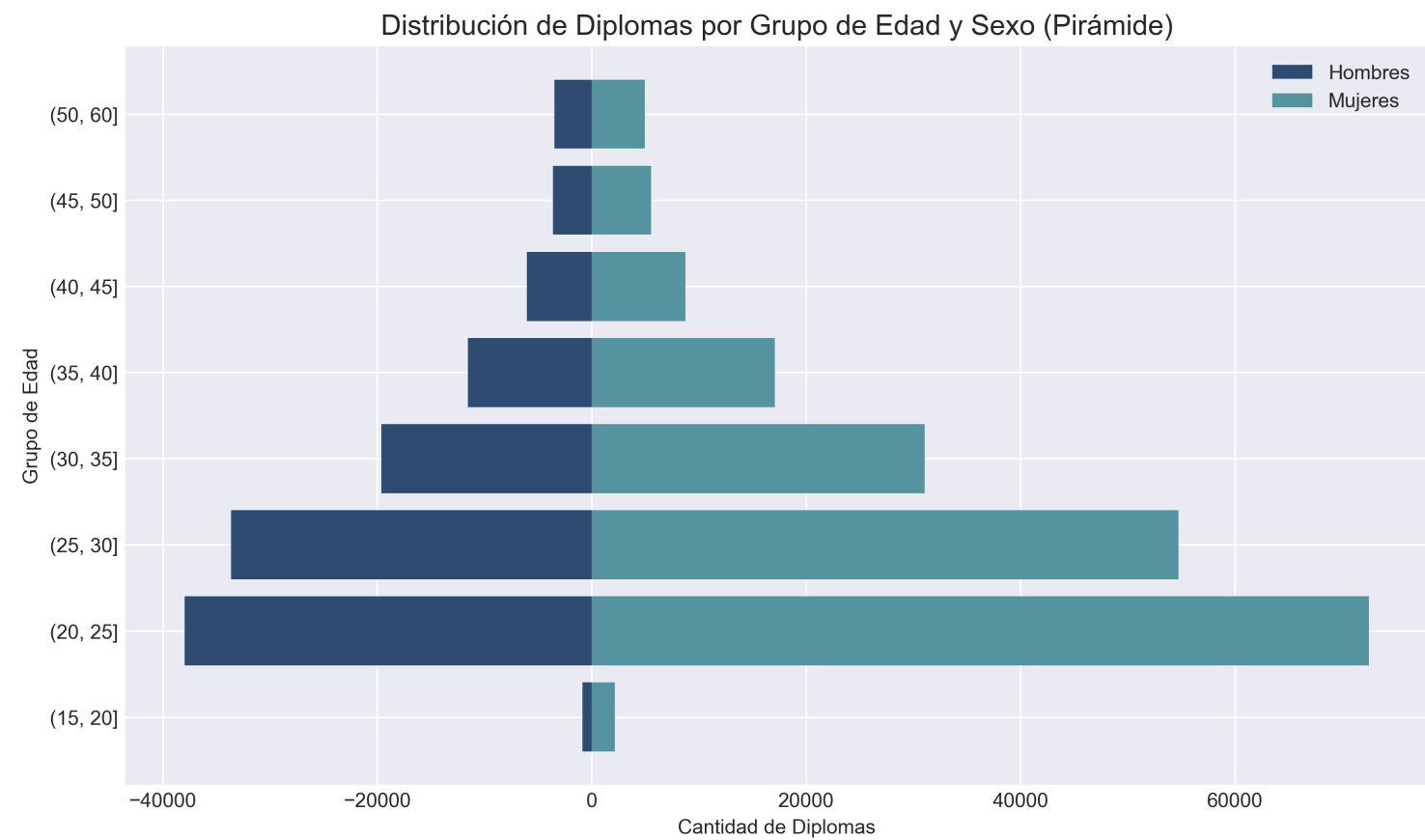


Diferencia constante de ~10,000 diplomas/año

```
df.groupby('Año').sum()['Cantidad']  
df.groupby('Año').sum()['Cantidad']  
df.groupby('Año').sum()['Cantidad']
```

6. Gráfico de Pirámide Poblacional

Definición: Barras horizontales bidireccionales que muestran dos grupos por intervalos.
Cuando usar: Análisis demográfico, estructura por edad y género.
Ejemplo: Distribución de diplomas por grupos de edad y sexo

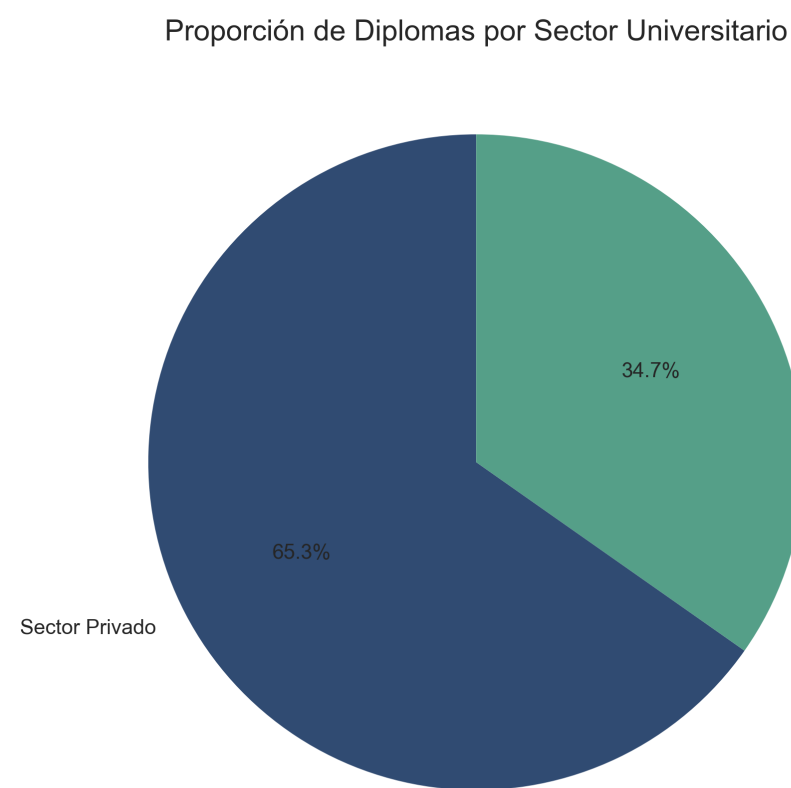


Mayor concentración: 20-30 años

```
df.groupby('Edad').sum()['Cantidad']  
df.groupby('Edad').sum()['Cantidad']  
df.groupby('Edad').sum()['Cantidad']
```

7. Gráfico de Pastel (Circular)

Definición: Círculo dividido en sectores proporcionales.
Cuando usar: Mostrar composición porcentual (2-5 categorías).
Ejemplo: Proporción diplomas sector estatal vs privado

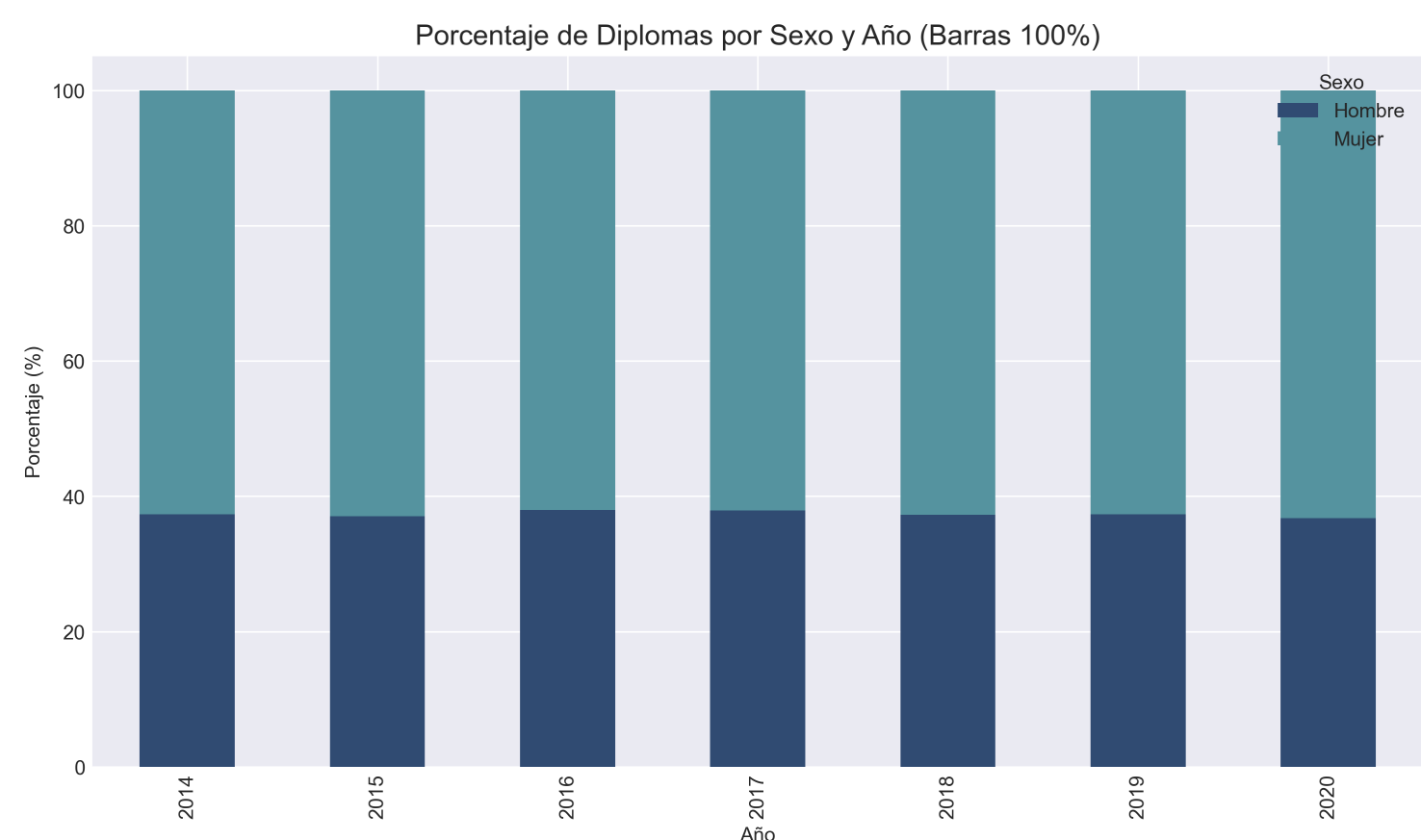


Privado: 65.3% | Estatal: 34.7%

```
df.groupby('Sector').sum()['Cantidad']  
df.groupby('Sector').sum()['Cantidad']  
df.groupby('Sector').sum()['Cantidad']
```

8. Gráfico de Barras 100 %

Definición: Barras normalizadas al 100% mostrando proporciones relativas.
Cuando usar: Comparar composiciones porcentuales.
Ejemplo: Evolución porcentual de género por año

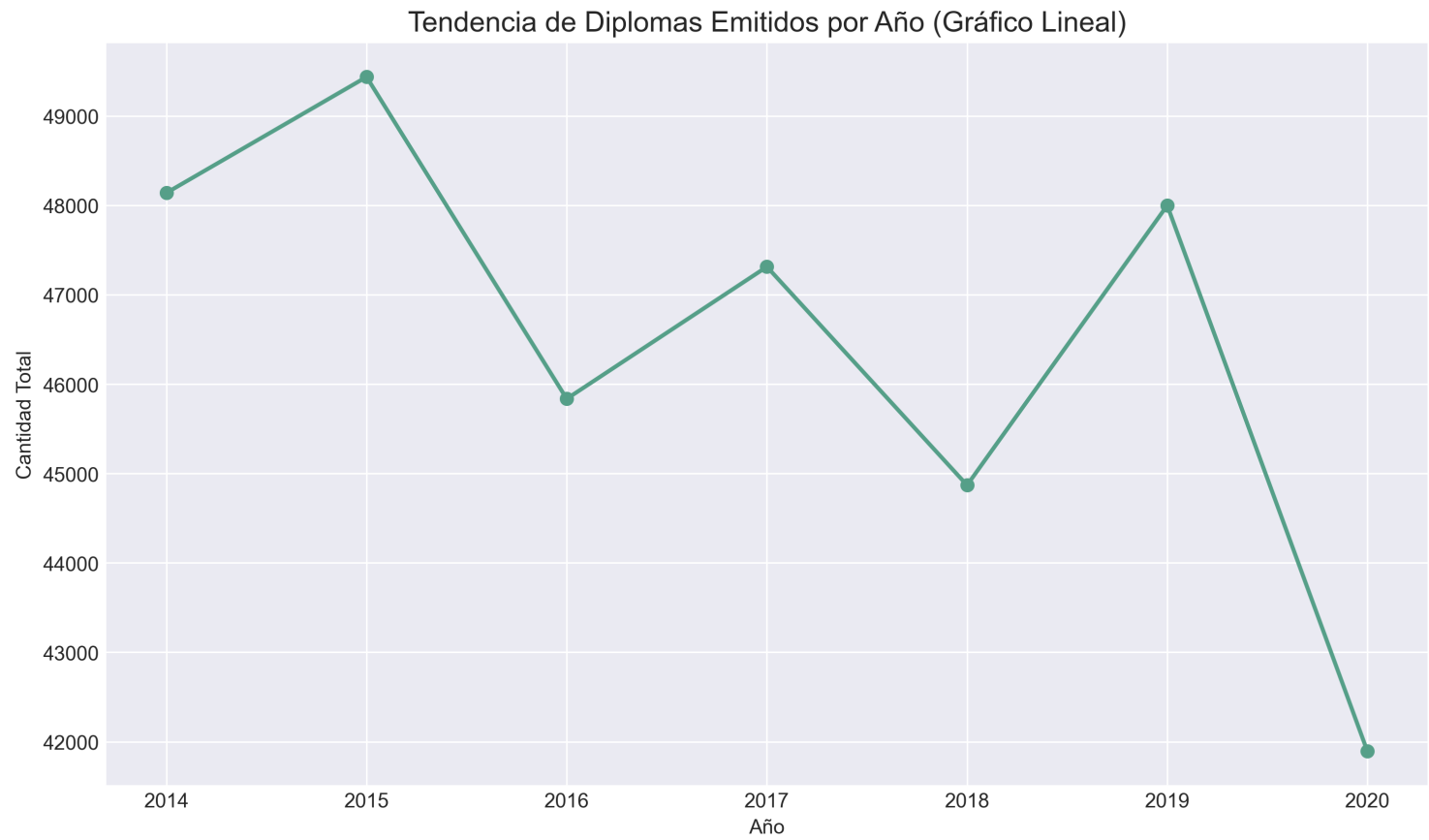


Proporción estable: 62% mujeres, 38% hombres

```
df.groupby('Año').sum()['Cantidad']  
df.groupby('Año').sum()['Cantidad']  
df.groupby('Año').sum()['Cantidad']
```

9. Gráfico Lineal

Definición: Línea continua que conecta puntos mostrando tendencias.
Cuándo usar: Series temporales o patrones de cambio.
Ejemplo: Tendencia de diplomas emitidos anualmente

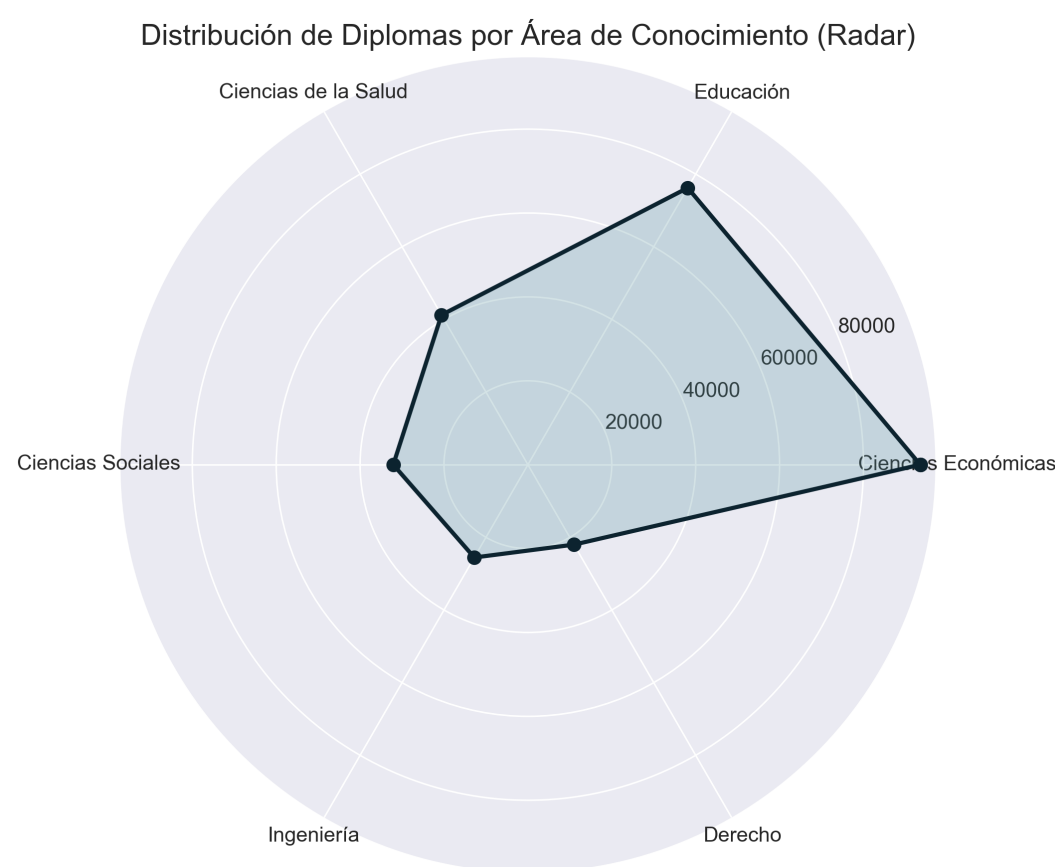


Tendencia irregular, caída en 2020

```
df = df[df['Año'] > 2010].reset_index(drop=True)
df['Cantidad'] = df['Cantidad'].rolling(window=3).mean()
df = df[df['Cantidad'] > 45000].reset_index(drop=True)
df.to_csv('diplomas_tendencia.png', index=False)
```

10. Gráfico de Araña (Radar)

Definición: Gráfico circular con ejes radiales.
Cuándo usar: Comparar múltiples variables simultáneamente.
Ejemplo: Distribución de diplomas por área de conocimiento



Educación lidera con > 80,000 diplomas

```
areas = ['Educación', 'Ciencias Económicas', 'Derecho', 'Ingeniería', 'Ciencias Sociales', 'Ciencias de la Salud']
max_val = 80000
df = df[df['Cantidad'] > 40000].reset_index(drop=True)
df.to_csv('diplomas_por_area.png', index=False)
```