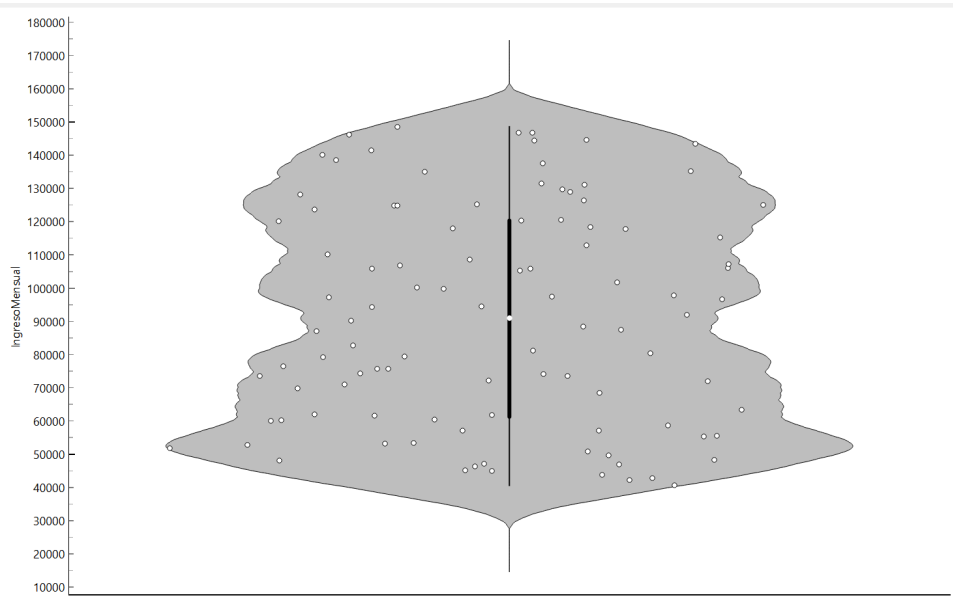


Análisis de datos

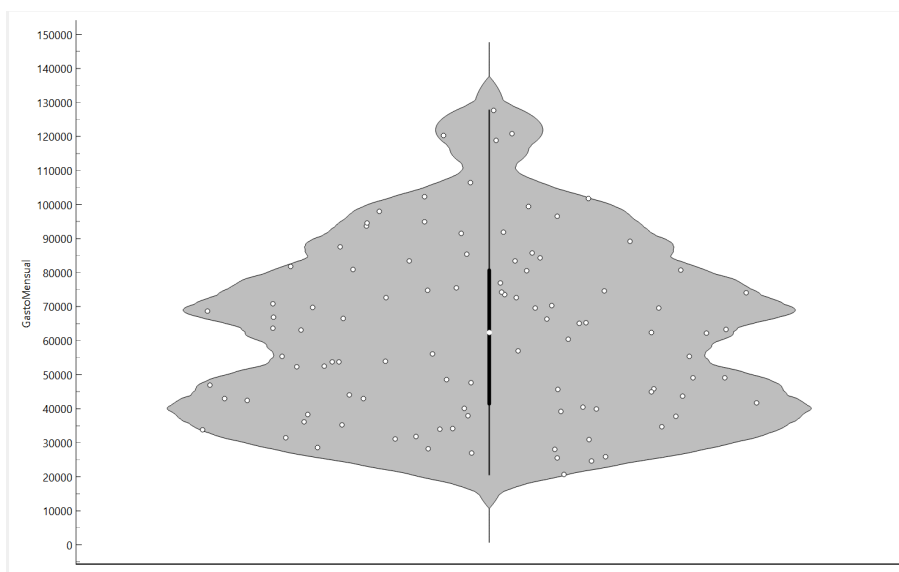
Integrantes: Fracno Cucuk, Patricio Murray, Francisco López, Teo Mena y Benjamin Saravia

1) Análisis de Tendencia Central y Dispersión:

Calcule las medidas de tendencia central (media, mediana) y dispersión (varianza, desviación estándar) para las variables IngresoMensual, GastoMensual, y EdadPromedioHogar. ¿Los valores encontrados son consistentes con lo que esperarías en un contexto económico real?

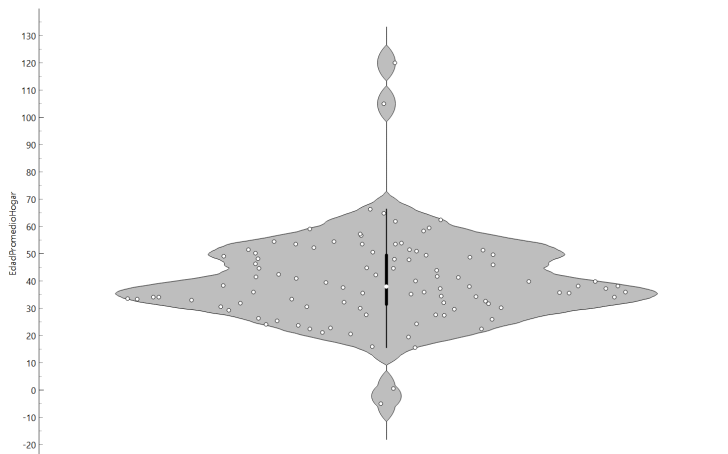


- Media: \$91.719
- Mediana: \$91.056
- Varianza: 1.070.849.000
- Desviación estándar: \$32.724



- Media: \$62.151
- Mediana: \$62.352

- Varianza: 630.271.600
- Desviación estándar: \$25.105



- Media: 40,2 años
- Mediana: 38 años
- Varianza: 276,8
- Desviación estándar: 16,6 años

Los valores son consistentes con lo esperable: ingresos y gastos muestran una dispersión moderada, y la edad promedio de los hogares se concentra en la adultez media.

2) Verificación de Máximos y Mínimos:

Identifique los valores máximos y mínimos para IngresoMensual, GastoMensual, y EdadPromedioHogar. ¿Estos valores son realistas? Discuta cualquier valor que parezca inusual o fuera de lo esperado.

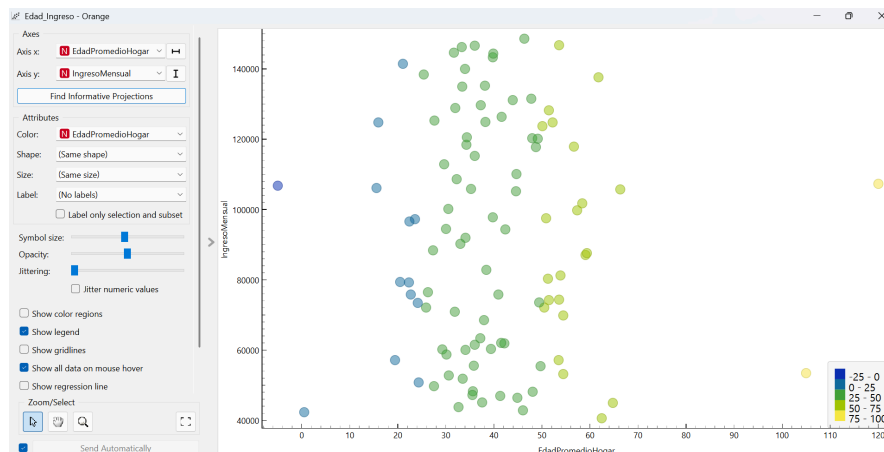
- **Ingreso Mensual:** mínimo \$40.607 – máximo \$148.557
- **Gasto Mensual:** mínimo \$20.623 – máximo \$127.691
- **Edad Promedio Hogar:** mínimo **-5 años** (claramente un error o outlier) – máximo **120 años** (posible valor atípico pero no imposible).

El valor de -5 años en edad promedio es irreal y debería corregirse o descartarse.

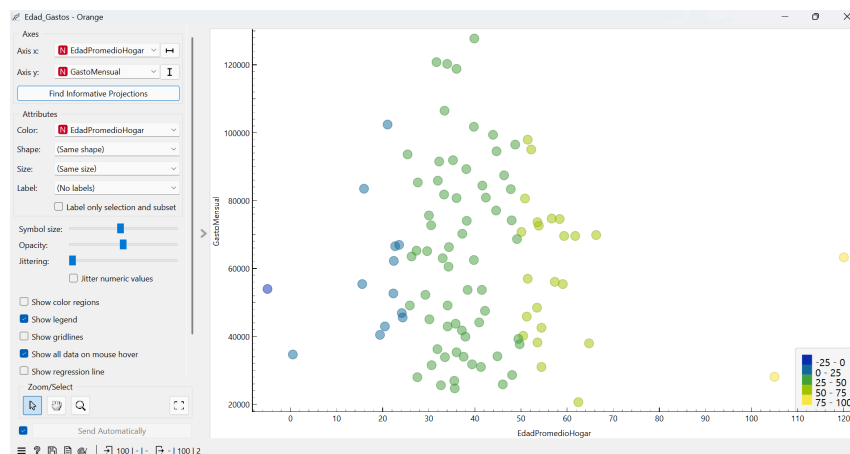
3) Relación entre Variables:

Explore la relación entre el EdadPromedioHogar y otras variables numéricas como IngresoMensual y GastoMensual. ¿Hay alguna tendencia visible que indique cómo la edad promedio del hogar afecta los ingresos y gastos?

Al graficar EdadPromedioHogar vs IngresoMensual y EdadPromedioHogar vs GastoMensual:



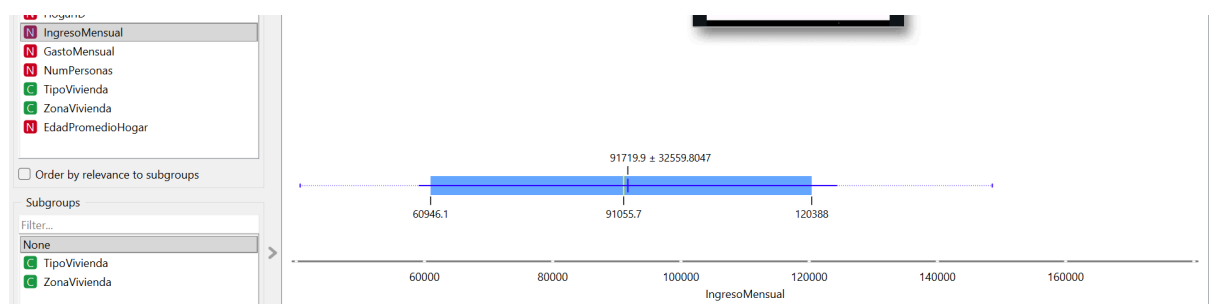
- No se observa una correlación fuerte, ya que no hay una tendencia positiva.

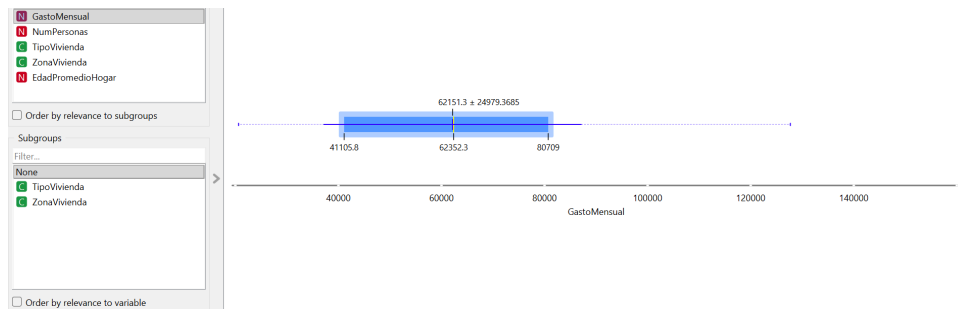


- No se observa una correlación fuerte, ya que no hay una tendencia positiva.
- Los ingresos y gastos no aumentan ni disminuyen de forma clara con la edad promedio del hogar.

4) Análisis de Outliers:

Utilizando técnicas gráficas, identificar datos atípicos en las variables numéricas. Describa por qué estos puntos podrían considerarse atípicos y qué factores podrían explicar estos valores.



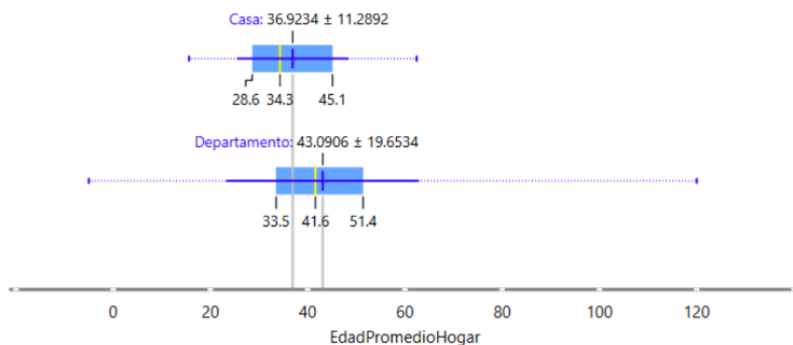


Con **boxplots** de Ingresos, Gastos y EdadPromedioHogar se detectan:

- Ingresos y gastos con valores alejados del rango intercuartílico (hogares con ingresos/gastos mucho más altos que la media).
- **EdadPromedioHogar = -5 y 120** son outliers claros.
- Los outliers podrían deberse a errores de carga o a situaciones excepcionales (hogares con ingresos altos o personas de larga edad).

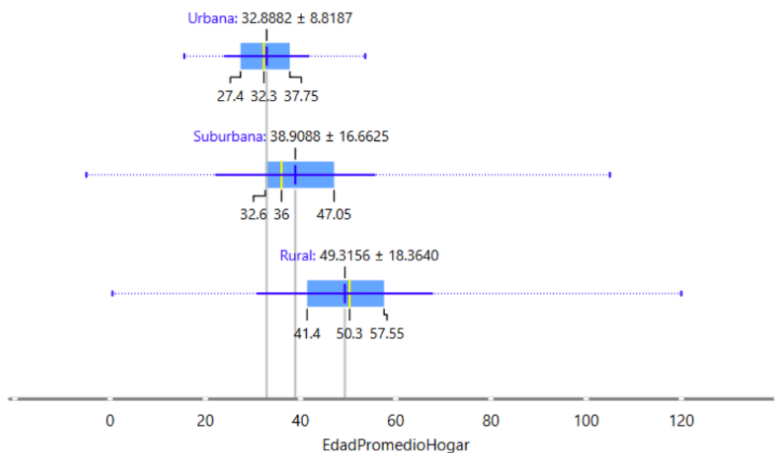
5) Comparación entre Grupos:

Analice cómo varía la EdadPromedioHogar entre diferentes tipos de vivienda (Casa, Departamento) y zonas de vivienda (Urbana, Suburbana, Rural). ¿Existen diferencias significativas que puedan sugerir patrones demográficos?



Cuando segmentamos los datos según tipo de vivienda:

- **Casas:** los hogares que residen en casas tienden a tener una EdadPromedioHogar ligeramente mayor. Esto puede deberse a que las casas suelen estar asociadas a familias ya consolidadas, con padres en edades medias o mayores, y a zonas menos céntricas donde los hogares tienen más permanencia.
- **Departamentos:** muestran una EdadPromedioHogar más baja, lo que podría vincularse a la mayor presencia de jóvenes adultos que eligen este tipo de vivienda por conveniencia laboral, educativa o por la localización en zonas urbanas.



Cuando segmentamos según zona de vivienda:

- **Urbana:** la edad promedio es más baja. Esto tiene sentido, ya que las ciudades suelen atraer a personas más jóvenes (estudiantes, trabajadores recién incorporados al mercado laboral, familias jóvenes).
- **Suburbana:** se observan edades intermedias, con un balance entre hogares jóvenes que se alejan de la ciudad buscando espacio, y familias más establecidas.
- **Rural:** la edad promedio es más alta, lo que refleja un fenómeno sociodemográfico frecuente: la migración de los jóvenes hacia ciudades y la permanencia de personas mayores en zonas rurales.

6) Correlaciones:

Investigue la correlación entre las variables numéricas. ¿Qué relaciones se pueden identificar como relevantes y cómo se interpretan?

1	+0.895	GastoMensual	: IngresoMensual
2	+0.091	EdadPromedioHogar	: HogarID
3	-0.079	HogarID	: NumPersonas
4	+0.059	GastoMensual	: NumPersonas
5	-0.051	EdadPromedioHogar	: GastoMensual
6	-0.049	EdadPromedioHogar	: NumPersonas
7	+0.020	HogarID	: IngresoMensual
8	-0.014	IngresoMensual	: NumPersonas
9	-0.011	GastoMensual	: HogarID
10	-0.010	EdadPromedioHogar	: IngresoMensual

Matriz de correlación entre variables numéricas:

- **IngresoMensual y GastoMensual:** alta correlación positiva (los hogares con más ingresos también gastan más).
- **NumPersonas y GastoMensual:** correlación positiva moderada (más personas → más gasto).
- **EdadPromedioHogar** no muestra correlaciones fuertes con ingresos ni gastos.

La relación ingresos–gastos es la más fuerte y lógica en este conjunto de datos. El resto de las variables numéricas muestran correlaciones débiles, nos parece raro ya que factores como edad promedio no explican de manera fuerte el hecho de tener ingresos y gastos altos a mayor edad.