Comentarios

1. Edad vs Severidad:
   1. Claramente la mayor cantidad de afectados en los accidentes se encuentra entre los 20 y los 40 años (ver histograma).
   2. El gráfico de barras y el de dispersión no me quedan claros.
2. Estado vs Edad por Sexo
   1. Si mi interpretación de los gráficos de caja y bigotes es correcta, podemos observar que:
      1. La distribución etaria de los ilesos es similar para ambos géneros. Yendo más lejos, me animaría a decir que esto puede obedecer a la “ley de los números grandes”, ya que esta categoría es la que más observaciones tiene
      2. Para la categoría “heridos”, noto que las distribuciones varían: los hombres jóvenes (hasta ~30 años) son seriamente afectados, y podríamos decir que el 75% de las observaciones para este género se comprende entre hombres de 0 a 40 años. Para las mujeres, la distribución entre el percentil 50 y el 75 es más amplia
      3. Finalmente, para la categoría “muertos”, la diferencia en la *agrupación* entre hombres y mujeres es aún más acentuada, especialmente de los percentiles 25 a 75. El 50% de los hombres que fallecen está entre 0 y 36 años aproximadamente, mientras que para las mujeres este rango es ligeramente mayor (de 0 a 43 años aprox.)
3. Estado vs Sexo:
   1. Algo destacado de este gráfico es notar cómo el número de mujeres heridas es mayor que el número de mujeres ilesas.
   2. Aún así, el número de mujeres heridas es menor que el de hombres (aproximadamente 20,000 observaciones menos)
4. Análisis por Modo:
   1. El número de muertos en automóvil parece ser cero (0), pero sospecho que se trata de un problema de escala.
   2. El número de heridos y muertos en bicicleta y motocicleta sí logra apreciarse. Además, para ambos modos, el número de heridos es mayor que el de ilesos.
   3. Me parece que definitivamente debemos agregar algunos modos para obtener datos “más representativos”. Propongo las categorías: automóvil (incluye campero y camioneta), Vehículo de Carga Pesada – VCP (incluye Volqueta y Tractocamión), Transporte Público (Bus, Buseta y Microbus).
5. **Estado vs Condición:**
   1. Con los conductores me parece que, aunque el gráfico da información intuitiva (hay un exageradamente mayor número de ilesos que de heridos y muertos debido a las condiciones de protección que ofrece el vehículo), el número de muertos puede estar sesgado en el *plot* debido a problemas de escala.
   2. El número de heridos en la condición “pasajero/acompañante” es llamativo. Una hipótesis podría ser que muchos de estos datos provienen de accidentes que incluyen motocicletas, donde el parrillero no tiene mayor protección; o que los acompañantes en los vehículos no utilizan el cinturón de seguridad.
   3. Si bien los usuarios vulnerables tienen un número de observaciones proporcionalmente bajo comparado con los conductores, me parece que hay un área de exploración importante allí. Es decir, un posible enfoque es plantear modelos solamente para usuarios vulnerables.
6. **Por LOCALIDAD**
   1. Creo que este análisis sería particularmente útil si lo vinculáramos con variables socioeconómicas. Me atrevería a decir que incluso sería útil para los modelos espaciales (versión 2.0 de esto :v)
7. **HEATMAPS**
   1. Por localidad: claramente hay un patrón en el número de observaciones (sobre todo para motocicletas y automóviles en las localidades que (asumo) son más grandes como Kennedy, Suba, Usaquén y Engativá.
   2. Curiosamente no encuentro un patrón claro entre los días de la semana. Esperaría que los fines de semana tuviéramos más observaciones. El plot no tiene los días de la semana organizados de forma convencional, pero puede entenderse fácilmente.
   3. Esperaría que hubiese una distribución *temporal* de los accidentes que indicara claramente las horas punta y horas valle. Aunque hay diferencias, no parecen ser representativas.