

PROCESO, JUSTIFICACIÓN Y EXPLICACIÓN DE LAS VISUALIZACIONES

En este documento se detalla para cada visualización:

- **Proceso** llevado a cabo para la extracción, tratamiento y preparación de los datos obtenidos de la fuente de datos abiertos seleccionada.
- **Justificación** de las visualizaciones estáticas realizadas como parte de la práctica.
- **Explicación** de los insights/conclusiones alcanzados y reflejados en las visualizaciones realizadas

El objetivo de incluir las visualizaciones en esta memoria no es que se puedan ver perfectamente al detalle (para ello están adjuntos los .png), es únicamente para tener una idea de a qué visualización se está haciendo referencia.

▪ Sobre el dataset

El dataset, que tiene 15 columnas en total, recoge datos de tiroteos de policías a civiles entre 2015 y 2020 en Estados Unidos. Para cada fila, es decir, cada tiroteo, se indican datos como:

- Fecha exacta
- Raza del asesinado
- Arma que portaba el asesinado
- Edad
- Género
- Ciudad del asesinato
- Estado del asesinato
- Signos de enfermedad mental

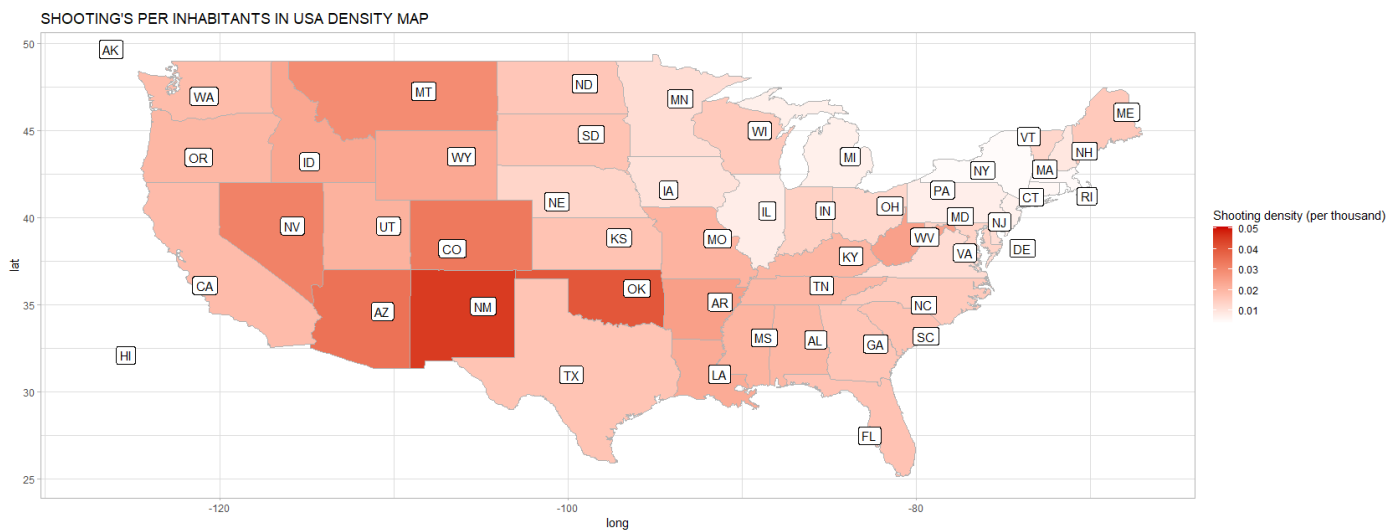
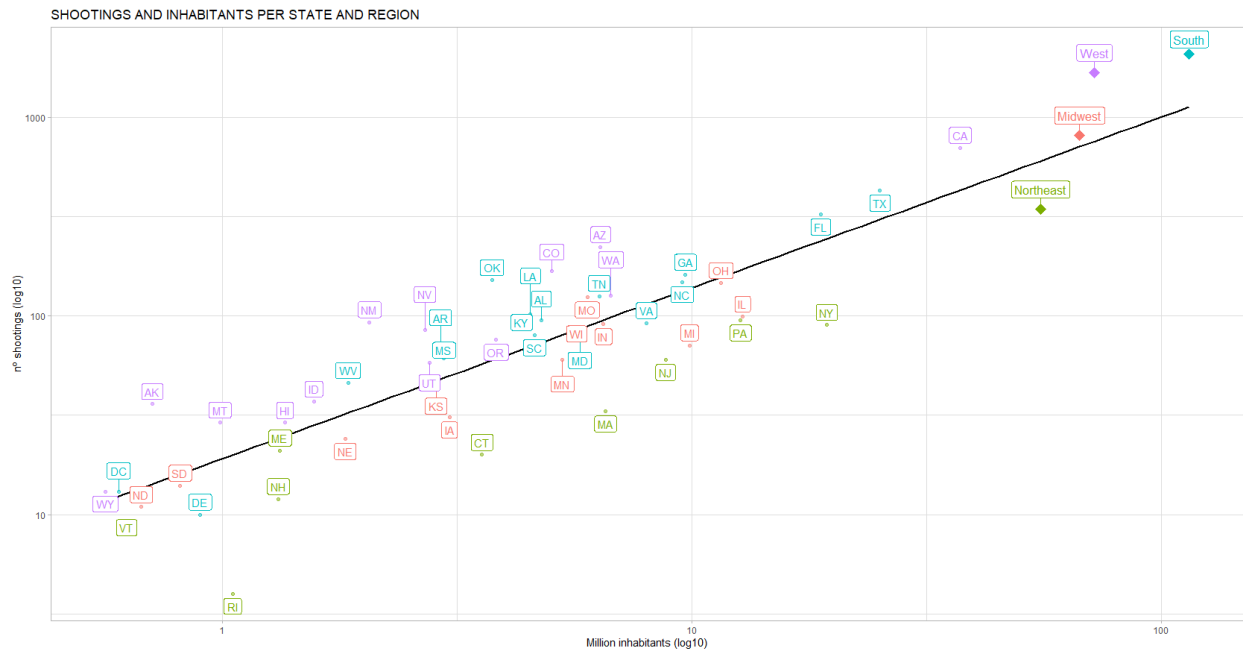
Estas 15 columnas han sido analizadas para ver qué columnas podían resultar realmente de interés para las visualizaciones y cuales no aportaban nada.

Se han realizado 7 visualizaciones que se dividen en 3 scripts:

- 2 visualizaciones geográficas: dónde y cuantos tiroteos ocurren
- 3 visualizaciones temporal-demográficas: cómo ha evolucionado este problema en el tiempo y cual es el perfil de los asesinados
- 2 visualizaciones de discriminación racial: ¿hay discriminación racial de los asesinados?, es decir, en el caso de ir desarmados ¿se asesinan más negros que blancos?

He optado por entregar los archivos en R en vez de Rmarkdown (en contra de lo que se recomendaba en el enunciado) ya que tenía todo el código comentado en los scripts de R. En la carpeta de soluciones también adjunto los png con las visualizaciones junto con los datasets que he usado y los scripts de R.

1 VISUALIZACIONES GEOGRÁFICAS



○ JUSTIFICACIÓN

Estás dos primeras visualizaciones ofrecen una perspectiva geográfica del problema de los tiroteos de policías a civiles en Estados Unidos.

○ PROCESO

Para la primera de las dos visualizaciones geográficas han sido necesarios dos datasets:

- Dataset de tiroteos de policías a civiles de 2015 a 2020 en EEUU
- Dataset censo de población por estados en 2010 (aunque sea de 5 años atrás suponemos que esta población no ha variado mucho)

El motivo por el que ha sido necesario el segundo dataset es porque los datos absolutos de tiroteos no ofrecen una perspectiva real de en qué estados el problema es mayor. Por ejemplo, California es el estado donde más tiroteos hay, pero a su vez es el estado más poblado de EEUU, sin embargo en estados como Texas, a pesar de que el número de tiroteos sea menor, el ratio tiroteos/habitantes es mucho mayor que en California.

El primer paso ha sido cruzar estos datasets por nombres de estados (haciendo la transcripción adecuada para que todo quedase bien). También se ha etiquetado aprovechando el dataset del censo, la región a la que pertenece cada estado (norte, sur, etc.). Con estas columnas y estos aesthetics se ha construido la primera visualización geográfica utilizando un `geom_point`:

- Nombre del estado: Etiquetas
- Habitantes: Eje x
- Número de tiroteos: Eje y
- Región: Color

Además, se ha superpuesto otro gráfico con los mismos aesthetics pero con el total de tiroteos de cada región. Los tiroteos totales por región están etiquetados con el mismo color que las regiones de cada estado por lo que estos puntos correspondientes a cada región hacen las veces de leyenda para la región de cada estado.

Para la segunda visualización se ha añadido una última columna al dataset llamada densidad de tiroteos que representa en tanto por mil el cociente entre el número de tiroteos en un estado y el número de habitantes. En este caso se ha utilizado un `geom_map` en el cual se han plasmado:

- Longitud: Eje x
- Latitud: Eje y
- Nombre del estado: Etiquetas
- Densidad de tiroteos: Relleno (escala continua de blanco a rojo)

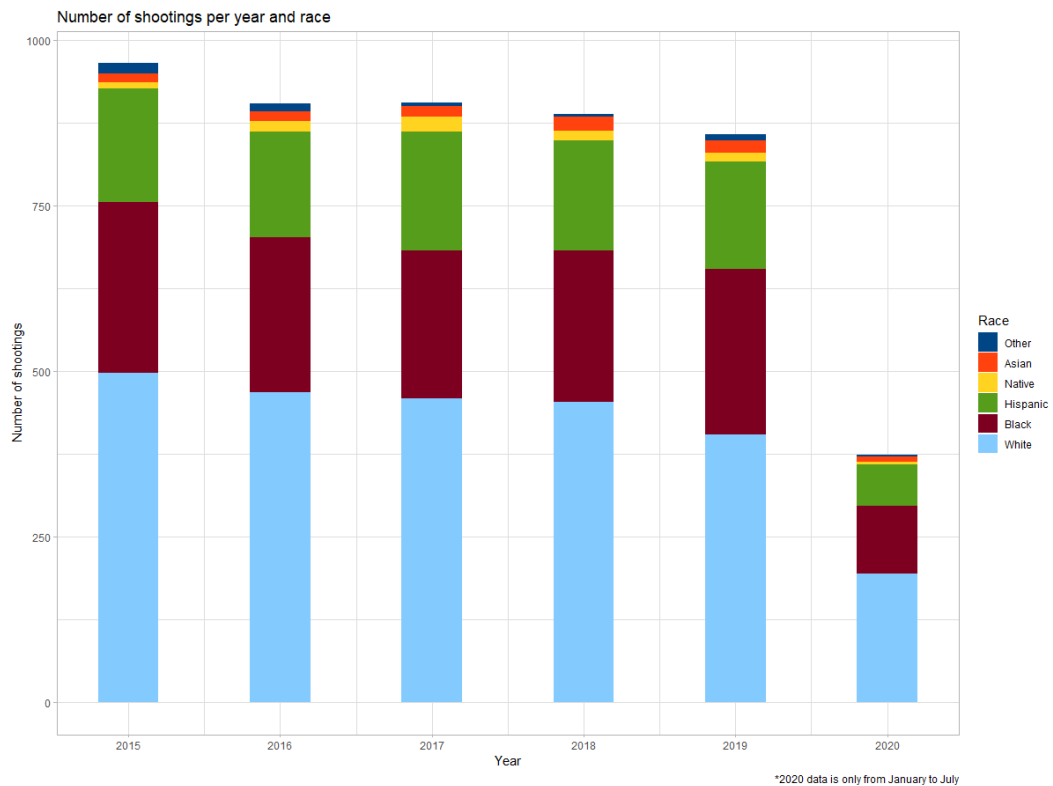
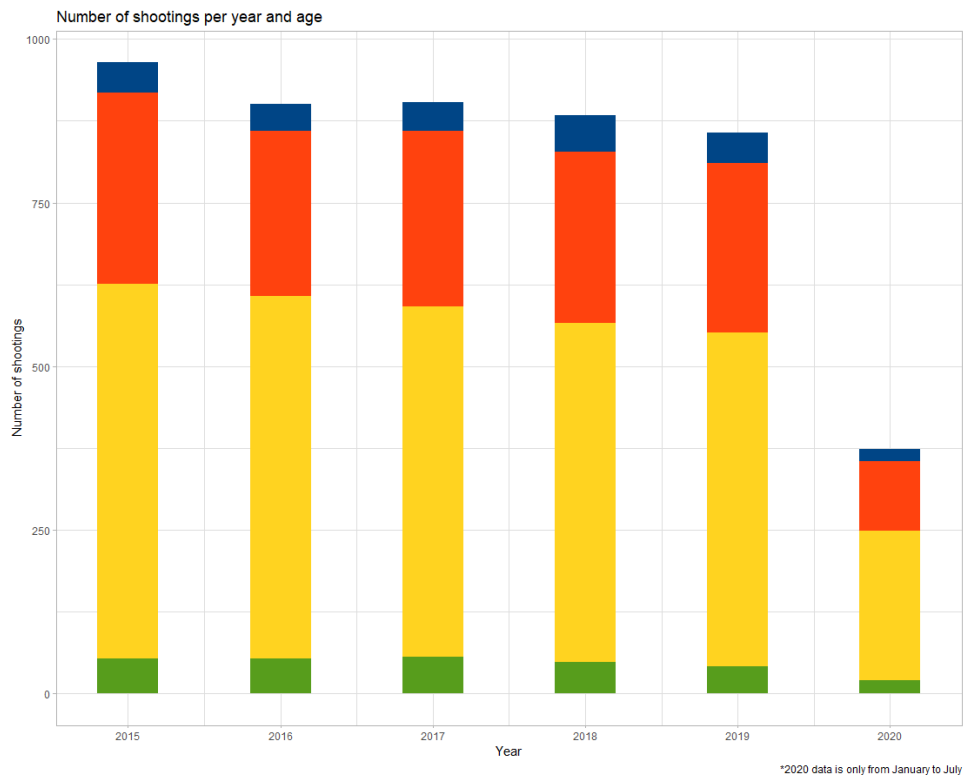
○ EXPLICACIÓN

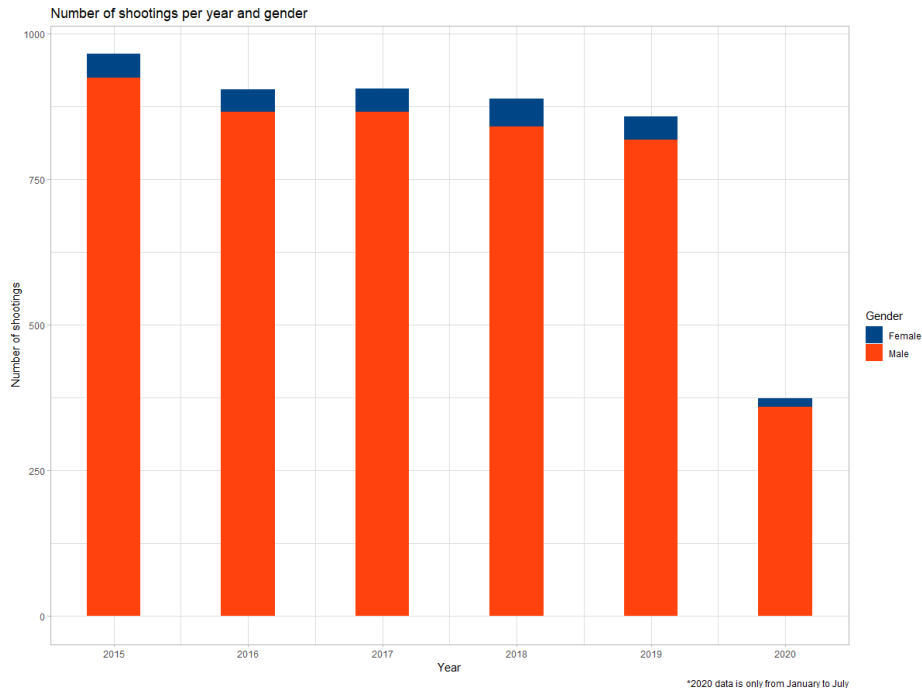
En estas dos visualizaciones podemos comprobar como el problema de los tiroteos de policías a civiles no se reparte de forma equitativa a lo largo y ancho de estados unidos. En el gráfico 1 vemos que la mayoría de estados por encima de la recta de regresión, es decir, aquellos cuya proporción de tiroteos por habitante es mayor, pertenecen a las regiones sur y oeste mientras que en las regiones medio-oeste

y noreste (especialmente esta última) tienen muchos menos tiroteos por habitante. En el mismo gráfico podemos obtener información como que California, Texas, Florida y Nueva York son los estados con mayor número de habitantes o que Rhode Island es un estado “outsider” en el que apenas hay tiroteos.

El gráfico 2 transmite información similar al gráfico 1 salvo que no podemos comparar los estados por número de habitantes o por número de tiroteos. A cambio podemos ubicar los estados geográficamente en el mapa. En este gráfico se corrobora que las zonas sur y oeste de Estados Unidos son las que mayor densidad de tiroteos tienen ya que están coloreadas de un rojo más intenso.

2 VISUALIZACIONES TEMPORAL-DEMOGRÁFICAS





○ JUSTIFICACIÓN

Las dos siguientes visualizaciones ofrecen una perspectiva demográfica y temporal de las personas que son tiroteadas por policías en Estados Unidos, es decir, a partir de estas visualizaciones podremos hacernos una idea del perfil medio de alguien que es tiroteado en Estados Unidos y cómo ha evolucionado a lo largo del tiempo este problema

○ PROCESO

El proceso para estas tres visualizaciones ha sido análogo por lo que se va a hacer una explicación general que puede aplicar a cualquiera de las tres variables que se analizan en estas visualizaciones: raza, edad y género.

Pasos seguidos:

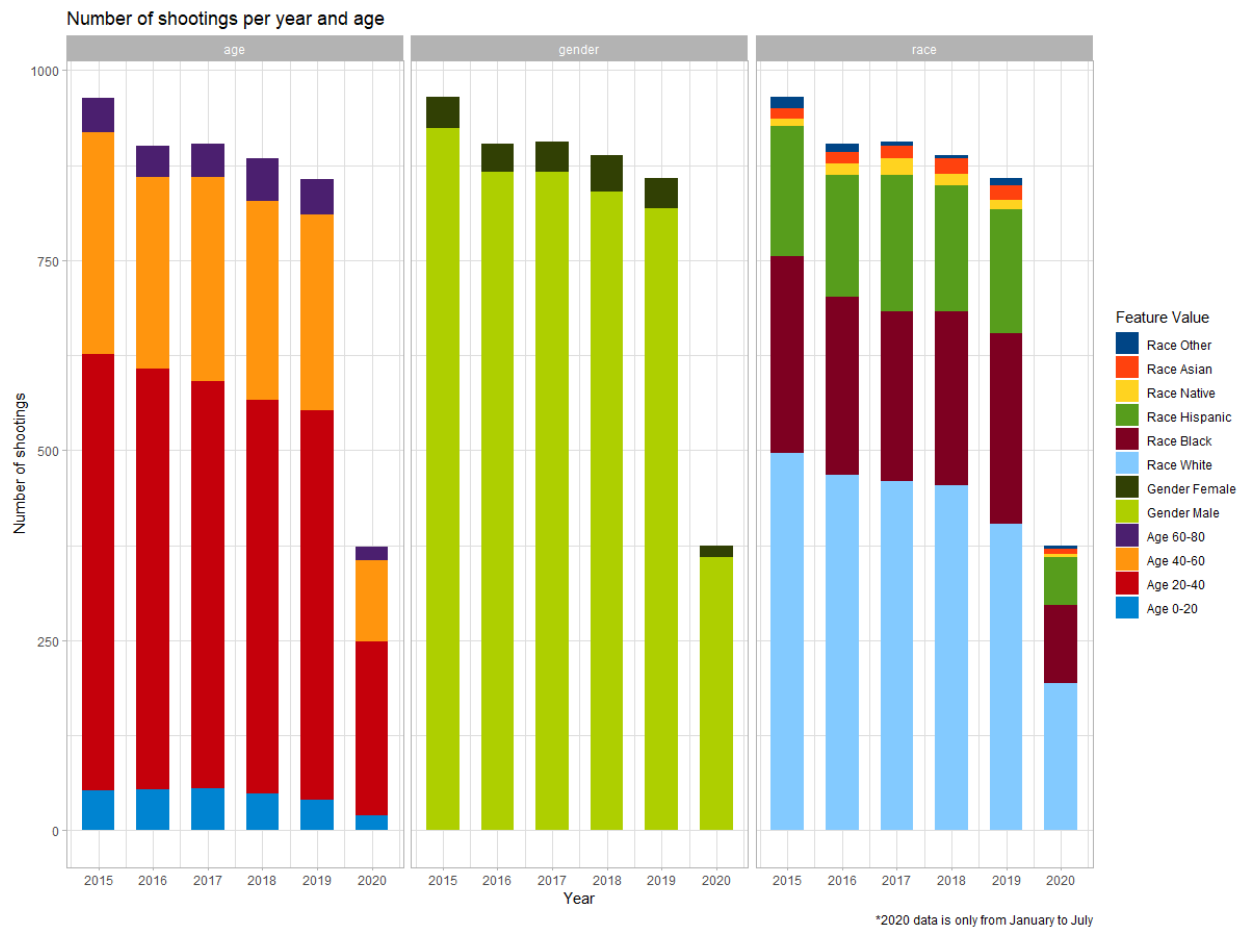
- Hacer one hot encoding de cada variable (en el caso de la edad haciendo uso de franjas de 20 años)
- Agrupar por año (habiendo separado primero el año de la fecha del tiroteo) y variable (raza, edad o género) y contar número de tiroteos
- Pivotar sobre la columna correspondiente a la variable que se va a visualizar (raza, edad o género).
- Factorizar la columna correspondiente a la variable para ordenar a mi gusto la leyenda (no tiene sentido que la raza “otros” sea la primera).

Hecho esto estamos listos para la visualización con una tabla en la que tengamos el número de tiroteos, la raza/edad/género y el año. Mandamos los aesthetics de la siguiente manera a un geom_col:

- Año: eje x

- Número de tiroteos: eje y
- Relleno: raza/edad/género

A modo de “bonus” se muestra a continuación una visualización en la que haciendo una peripecia he conseguido hacer facetting de las variables que estaba analizando (raza, edad y género). No obstante, no me gusta como queda esta visualización y el facetting está pensado para que el aesthetic “fill” sea el mismo en las distintas tablas, no que en cada tabla los colores sean distintos (por eso en la leyenda está todo junto sin separar por variables).



○ EXPLICACIÓN

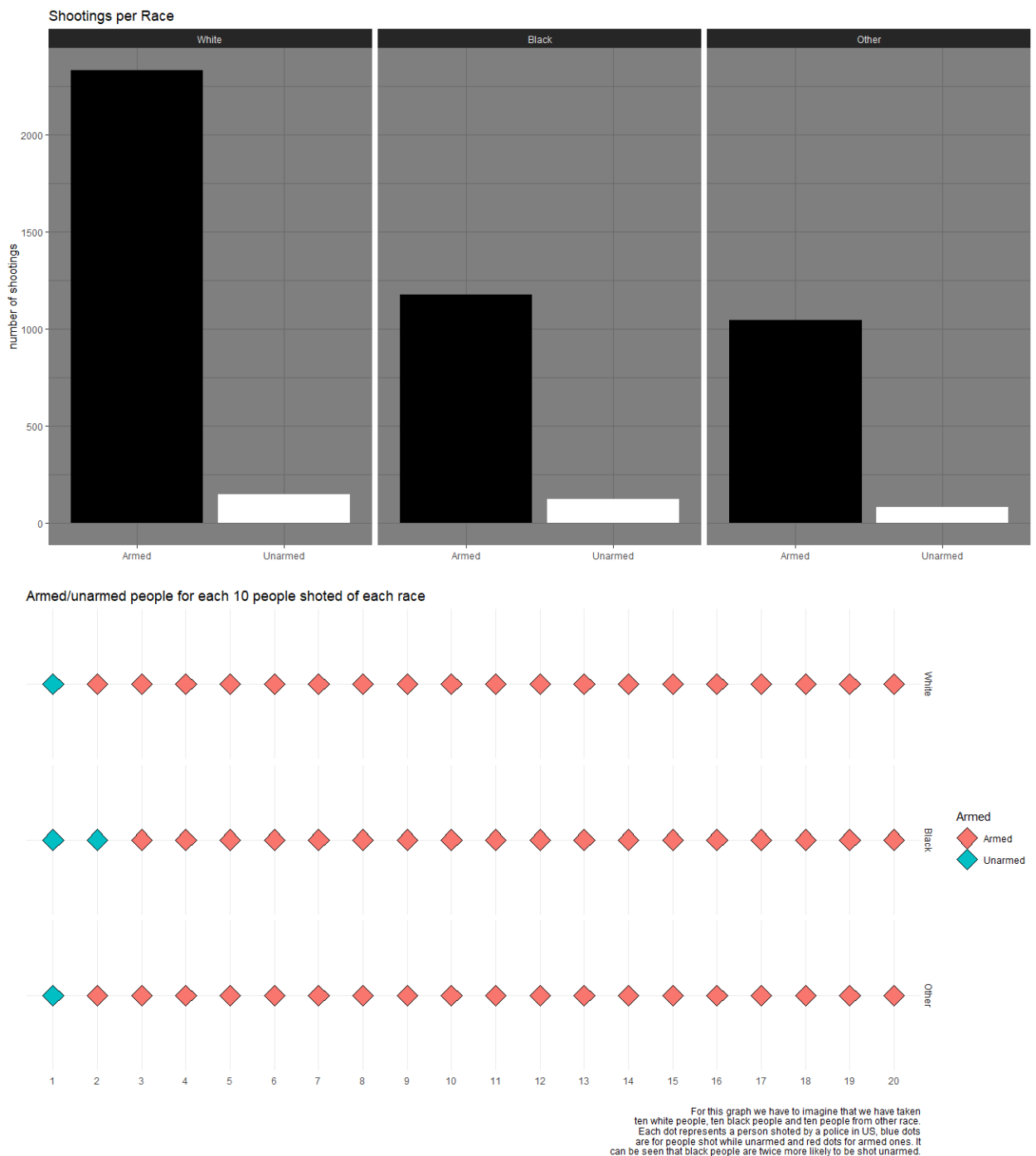
De estas tres visualizaciones se pueden sacar dos conclusiones claras:

- Dado que el número de asesinatos a lo largo de los últimos 5 años se mantiene prácticamente igual (hay que tener en cuenta como se indica al pie del gráfico que los datos de 2020 son solo hasta Junio), podemos concluir que el problema de los tiroteos de policías a civiles no ha mejorado significativamente desde hace 5 años aunque sí presenta una tendencia ligeramente descendente
- Si nos fijamos en los colores del gráfico de barras podemos ver el perfil medio de los asesinados: varones de raza blanca, negra o hispanica de entre 20 y 60 años. Hay que puntualizar, que para ir un paso más allá, habría que examinar y comparar exhaustivamente estos datos con el perfil

de ciudadano medio estadounidense y el perfil de delincuente medio estadounidense. Algunas conclusiones claras que se pueden sacar sin mayor indagación son:

- Los varones se ven envueltos en tiroteos muchas más veces que las que deberían si nos atenemos al censo de varones y mujeres en estados unidos (que uno puede intuir que no distará mucho de mitad y mitad).
- De la misma manera que las mujeres, los niños y los ancianos no se ven envueltos en muchos tiroteos comparado con la proporción que intuitivamente (sin ponerse a mirar censos) representarán en la sociedad norteamericana.

3 VISUALIZACIONES DISCRIMINACIÓN RACIAL



○ JUSTIFICACIÓN

El objetivo de las dos últimas visualizaciones es ilustrar si existe una discriminación racial por parte de los policías a la hora de disparar a civiles. Para ver esto hay que comparar por raza el número de civiles asesinados mientras iban armados y mientras iban desarmados.

○ PROCESO

Estas dos visualizaciones tienen procesos distintos por lo que se explicarán por separado.

▪ Visualización 1

Pasos:

- Seleccionamos las columnas que vamos a necesitar (raza y arma).
- En raza transformamos la columna para solo tener “blanco”, “negro” y “otro”. De la misma manera transformamos la columna arma para tener “armado” y “no armado”
- Agrupamos por raza y armado/no armado
- Factorizamos y ponemos niveles a los valores de las columnas para ordenar al gusto en la visualización
- Hacemos la visualización con los siguientes aesthetics:
 - Número de tiroteos: eje x
 - Armado/desarmado: eje y
 - Raza: Faceting

▪ Visualización 2

Esta visualización es completamente distinta a las otras ya que no se ha construido modificando el dataset mediante pivotes, agrupaciones, etc. Sino que se ha construido artificialmente. Para esta visualización, utilizando R he calculado

- Porcentaje de blancos asesinados desarmados respecto de blancos asesinados: 5,9%
- Porcentaje de negros asesinados desarmados respecto de negros asesinados: 9,5%
- Porcentaje de otras razas asesinados desarmados respecto de otras razas asesinados: 7,0%

Para mayor impacto de esta visualización he redondeado a lo bruto para obtener números enteros (a continuación, se verá por qué). Los redondeos han sido:

- Porcentaje de blancos asesinados desarmados respecto de blancos asesinados: 5%
- Porcentaje de negros asesinados desarmados respecto de negros asesinados: 10%
- Porcentaje de otras razas asesinados desarmados respecto de otras razas asesinados: 5%

De esta manera podemos afirmar que:

- De cada 20 blancos asesinados 1 va desarmado
- De cada 20 negros asesinados 2 van desarmados
- De cada 20 otras razas asesinados 1 va desarmado

Estos números redondos ayudarán a que la visualización tenga más impacto.

El objetivo de la visualización es para cada raza pintar veinte figuritas (cada una representa una persona) y pintar de rojos las personas armadas y de azul las desarmadas. Para ello construyo un dataset con tres columnas: raza (eje y), armado/desarmado (color de relleno) y una tercera columna auxiliar para mandarla al eje x (que no significa nada).

○ EXPLICACIÓN

▪ Visualización 1

De esta visualización podemos obtener dos conclusiones

- Se asesina aproximadamente el doble de blancos que de negros. Esta proporción no se corresponde con la proporción racial en Estados Unidos (aproximadamente 60% blancos, 14% negros). Pero probablemente sean mucho más relevantes que el censo total otras medidas como la proporción racial de los delincuentes o la proporción racial de las personas con licencia de armas a la hora de sacar conclusiones.
- Una conclusión indudable de este gráfico es que independientemente de la raza la inmensa mayoría de los asesinados iban armados (lo cual liga el problema de los tiroteos de policías a civiles con el problema de las licencias de armas en Estados Unidos).

▪ Visualización 2

Esta visualización está hecha con el único propósito de arrojar una conclusión. Las personas de raza negra tienen el doble de probabilidad de ser asesinadas desarmadas por un policía. No obstante, esta visualización contextualiza este “doble de probabilidad” ya que no es lo mismo el doble de probabilidad de que te toque la lotería (que sigue siendo muy cercano a 0) que el doble de probabilidad de que salga cruz al tirar una moneda (que quiere decir que las dos caras de la moneda son cruces). En concreto de cada 20 blancos asesinados 1 va desarmado, mientras que de cada 20 personas negras asesinadas 2 van desarmados, esto es, es un 5% más probable ser asesinado desarmado si tu raza es negra.