**Observaciones:**

* **Todo lo realizado debe ser justificado de manera clara y precisa, con argumentos estadísticos, sino no se le dará valoración.**
* **La actividad puede ser entregada en parejas**
* **Cualquier intento de plagio y/o fraude será sancionado con una clasificación de 0.0**

**Análisis previo de datos (Grupo1)**

**Parte 1. Análisis de valores atípicos**

1. Para realizar la Actividad 1, debes crear un Rscript en RStudio y guardar el archivo con un nombre específico de tu preferencia.
2. Instala el paquete “carData” y carga los datos llamados “States”
3. Describe el dataset y cada una de sus variables
4. Realiza un summary de la variable dollars e interpreta dos estadísticas.
5. Para la variable dollars realiza un análisis de datos atípicos y realiza la depuración completa de la variable. Debes ir copiando todos los resultados que vayas obteniendo en R e ir interpretando y/o explicando tus procedimientos.
6. Realiza un summary de la variable dollars e interpreta dos estadísticas (las mismas del punto 4) de cada variable y compara con los resultados del punto 4
7. Revisa el supuesto de normalidad para la variable dollars. Explica e interpreta los resultados que encuentras. Debes mostrar por lo menos un análisis gráfico y una prueba o test estadístico.

**Parte 2. Análisis de valores ausentes**

1. Carga los datos llamados “attenu”
2. Describe el dataset y cada una de sus variables
3. Realiza un análisis de valores ausentes:

* menciona si hay variables con valores ausentes,
* describe la (s) variable(s) que hayas encontrado con valores ausentes,
* realiza la imputación de la misma y realiza un summary e interpreta
* Realiza una comparativa de las estadísticas que obtienes del summary de la variable omitiendo los valores ausentes y después de imputarlos.

**SUBE A SAVI EL DOCUMENTO EN WORD CON TUS RESULTADOS E INTERPRETACIONES**

**SUBE A SAVI EL ARCHIVO DE R DONDE DEJASTE TU CODIGO**

**Descripción del data frame**

“States” está estructurada por 51 filas y 8 columnas. Las observaciones están conformadas por los estados de EE. UU. y Washington, D. C. la información fue obtenida de *“United States (1992) Statistical Abstract of the United States. Bureau of the Census.”*

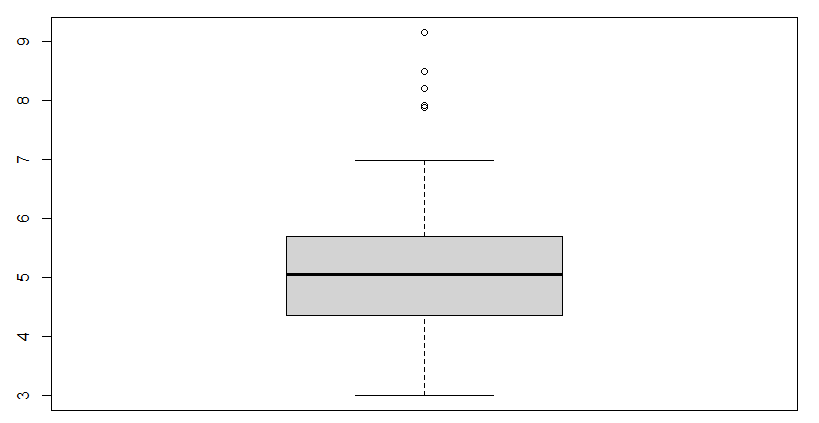
Las columnas del data frame son:

* Región – (Factor)
* Pop – (int)
* SATV – (int)
* SATM – (int)
* Percent – (int)
* Dollars – (num)
* Pay – (int)

**Interpretación de la variable dollars**

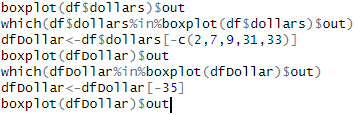
* La variable dollar cuenta con un valor mínimo de 2.993, esto quiere decir que es el valor más pequeño de esta variable.
* La variable dollar cuenta con un valor máximo de 9.159, esto quiere decir que es el valor más grande de esta variable.

**Análisis de datos atípicos**

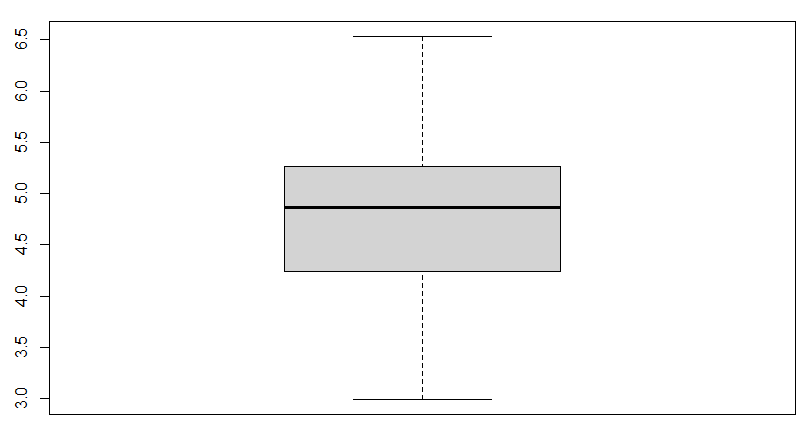


Los valores atípicos encontrados fueron 7.887, 7.914, 8.210, 9.159, 8.500, en las posiciones 2, 7, 9, 31, 33 respectivamente

Proceso de la depuración:



Resultado de la depuración:



**Interpretación estadística después de la depuración:**

Estadísticas del data frame sin depurar:

* Antes de depurar:

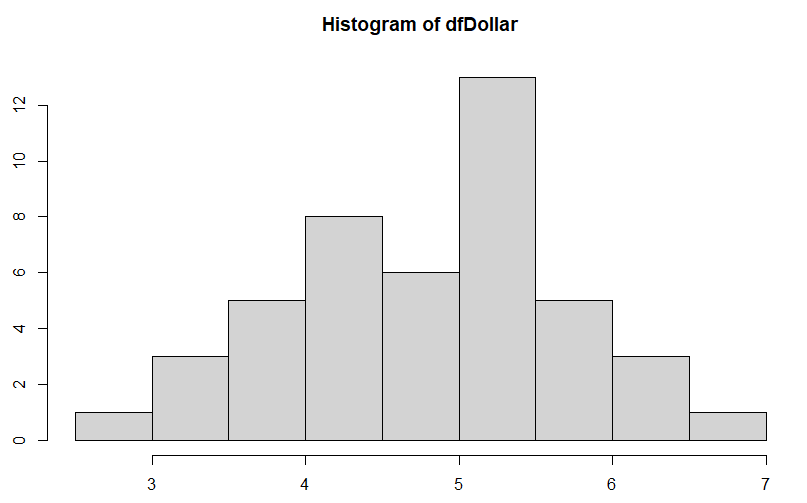
* Después de depurar:

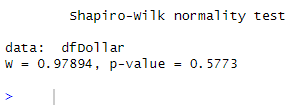


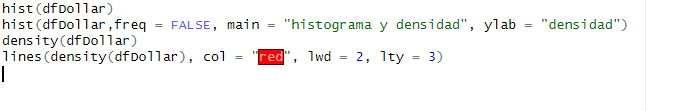
Como se puede ver el mínimo no se ve afectado después de la depuración, en cambio como entre los datos atípicos se encontraba el valor máximo este cambio de 9.159 por un valor de 6.534.

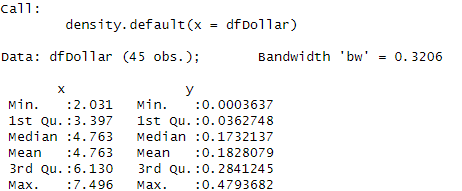
**Supuesto de normalidad:**

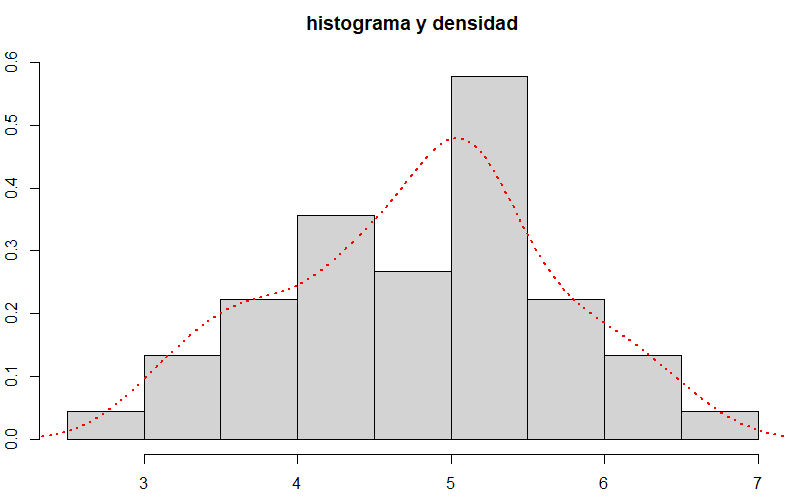
****

****

****

****

****

****