

# Tarea 3

*Melany Waldisperg*

*16 de septiembre del 2018*

## Introducción

Durante este estudio se realizó un análisis de la base de datos del programa “Mussel Watch”, disponible en la página web [www.kaggle.com](http://www.kaggle.com). El programa “Mussel Watch” consiste en el monitoreo de contaminantes en bivalvos para Estados Unidos desde 1986. Este programa tiene el fin de reunir información para evaluar el estado ambiental, mediante el análisis del taxa bivalvos conocidos mundialmente como bioindicadores (Hédouin et al. 2011; Waykar and Deshmukh 2012). En este contexto, los objetivos de este trabajo son determinar si el tamaño y estado de los bivalvos se correlaciona con el sexo, determinar las localidades más contaminadas y evidenciar si el estado de los bivalvos se correlaciona con el estado ambiental. Además, se espera encontrar bivalvos en peor estado, en localidades más contaminadas y durante años con mayor número de contaminantes.

## Metodología

Para poner a prueba nuestra hipótesis se seleccionaron 9 variables del listado disponible en la base de datos descargada, que incluyeron: especie, localidad, año, sexo, largo de la valva, atrofia tubular digestiva, tipo de contaminante, valor del contaminante y unidad de medida del contaminante. Posteriormente, se seleccionaron los años con mayor cantidad de datos disponibles (1995-1998) y se clasificaron los contaminantes en 3 grupos según su unidad de medida (ng/dry g, ng/dry g o L y ng Sn/dry g). Una vez filtrada la información seleccionada se terminó trabajando con información de una sola especie de bivalvo, denominado en la base de datos como “Dreissena”. Las variables descriptivas del estado de los bivalvos fueron largo de la valva y atrofia tubular digestiva, mientras que, la variable descriptiva del estado ambiental fue la concentración de contaminantes. Por último, se realizó el análisis de los datos en RStudio con los paquetes “tidyverse”, “knitr”, “readr”, “broom” y “rmarkdown” (Wickham 2017; Wickham, Hester, and Francois 2017; R Core Team 2018; Robinson and Hayes 2018; Xie 2018; Xie, Allaire, and Golemund 2018)

## Resultados y discusión

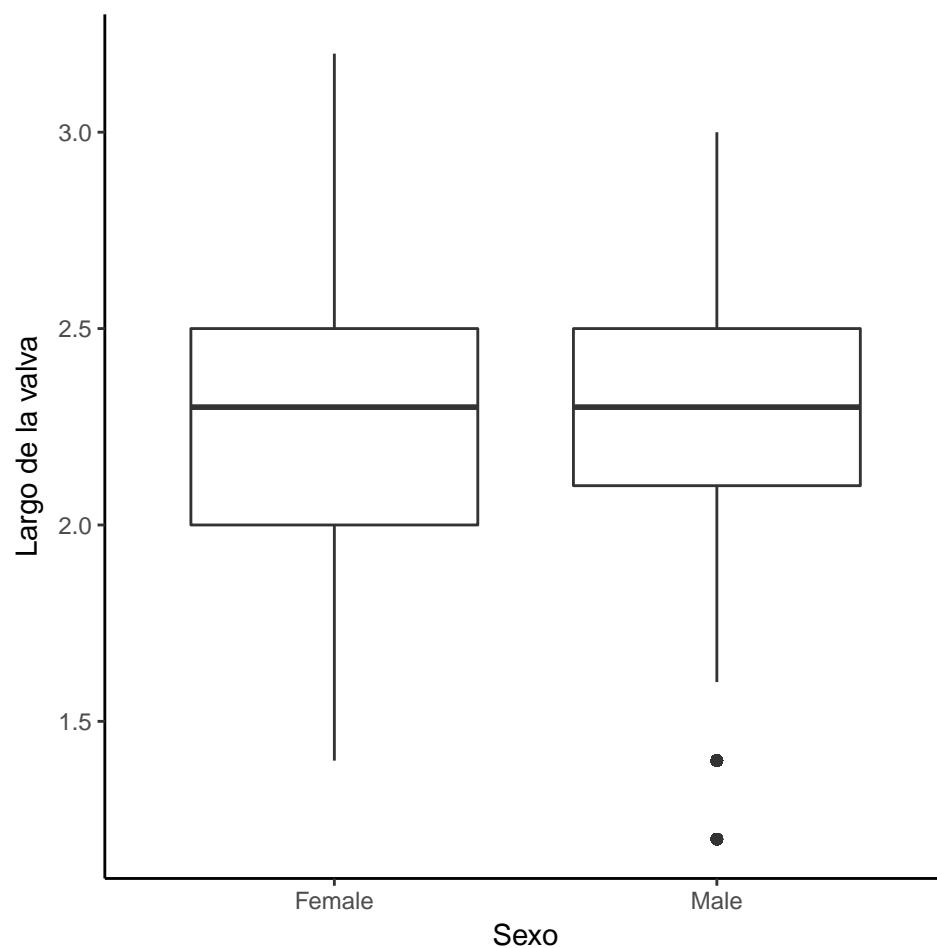
Con el fin de evidenciar diferencias potenciales en la captación de contaminantes por sexo explicadas por diferencias de tamaño, se realizó un analisis comparativo donde no se obtuvo diferencias significativas entre hembras y machos. El largo promedio fue igual a 2.26, 2.27 y la desviación estándar fue de 0.38, 0.36, para hembras y machos respectivamente (Fig. 1). Además, el modelo con mejor ajuste para explicar el largo de la valva contenía el sexo de los ejemplares como única variable explicativa (Tabla 1). Al evaluar la existencia de diferencias en el estado de los ejemplares según el sexo, no se encontraron diferencias estadísticamente significativas. La atrofia tubular digestiva promedio fue igual a 1.82, 1.85 y la desviación estándar fue de 0.76, 0.85, para hembras y machos respectivamente (Fig. 2). Nuevamente, el modelo con mejor ajuste para explicar, en este caso, la atrofia tubular digestiva contenía el sexo de los ejemplares como única variable explicativa (Tabla 2). Por otro lado, las localidades con mayor cantidad de contaminantes fueron el Lago Michigan para los tres grupos de contaminantes y ,además, el Lago Erie para el grupo contaminante 3 (Fig. 4, 6 y 8). Sin embargo, al comparar las gráficas de las concentraciones de contaminantes (Fig. 4, 6 y 8) con las de valores de atrofia tubular digestiva (Fig. 3, 5 y 7) no se evidenció una correlación del estado de los bivalvos con el estado ambiental. Por el contrario, el modelo obtenido con mejor ajuste a la concentración de contaminantes fue aquel que contenía la atrofia tubular digestiva de los ejemplares como única variable explicativa (Tabla 3, Fig. 9). Según la gráfica de la predicción de este último modelo se esperaría un incremento en los niveles de contaminante ambiental al aumentar la atrofia tubular digestiva (fig. 9). Sin embargo, hay que tener cuidado al analizar estos últimos resultados entregados por el modelo, ya que, no se están tomando en consideración los años y las localidades que parecen incrementar la incertidumbre que se ve reflejada en las figuras 3, 4, 5, 6, 7 y 8. Además, la falta de información para ciertos años y localidades podría estar alterando los resultados.

## Conclusión

El análisis realizado en el presente estudio con la base de datos “Mussel Watch” no nos permitió sustentar la hipótesis planteada, producto de diferencias en los resultados obtenidos mediante diferentes metodologías de análisis de datos.

## Anexos

### Figuras



### Tablas

Table 1: Tabla 1. Indices de ajuste de los modelos de medición del largo de la valva.

r.squared	AIC	Modelo	DeltaAIC
0.2416104	12359.41	Fit3	0.00000
0.2416104	12359.41	Fit4	0.00000
0.2406706	12383.89	Fit1	24.47042
0.0002777	18262.63	Fit2	5903.21494

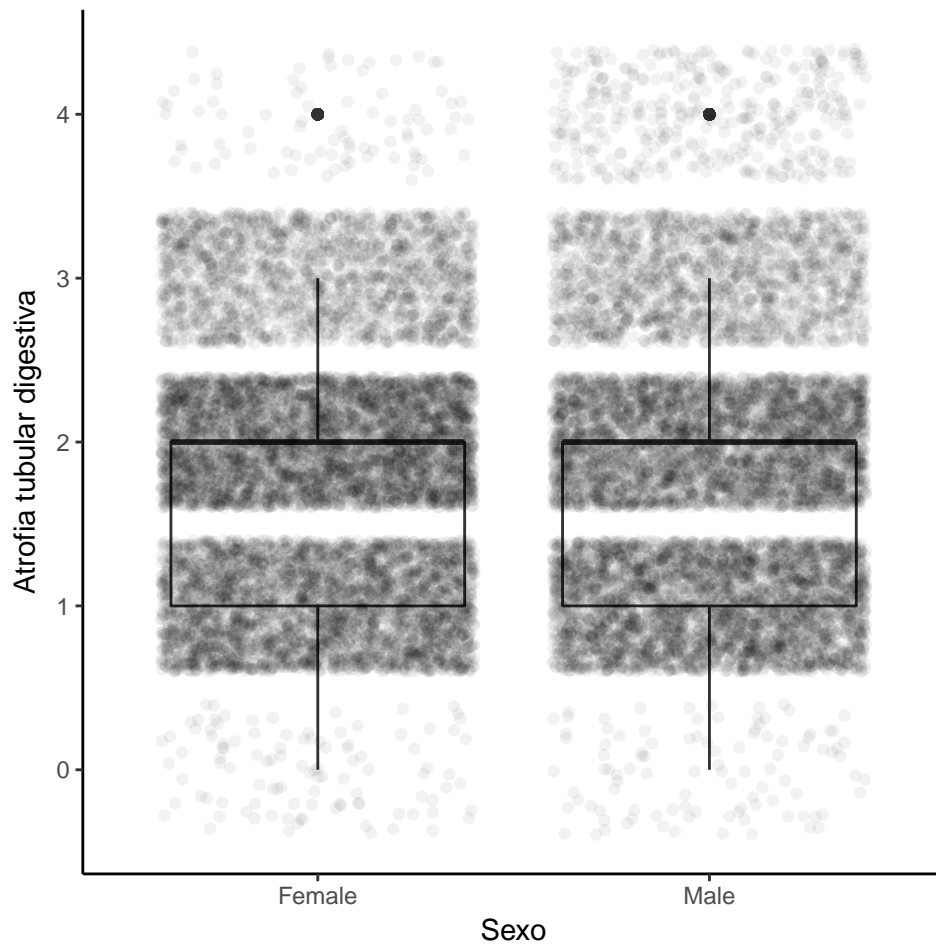


Figure 1: Figura 2. Gráfico de cajas con dispersión de puntos de la atrofia tubular digestiva vs el sexo.

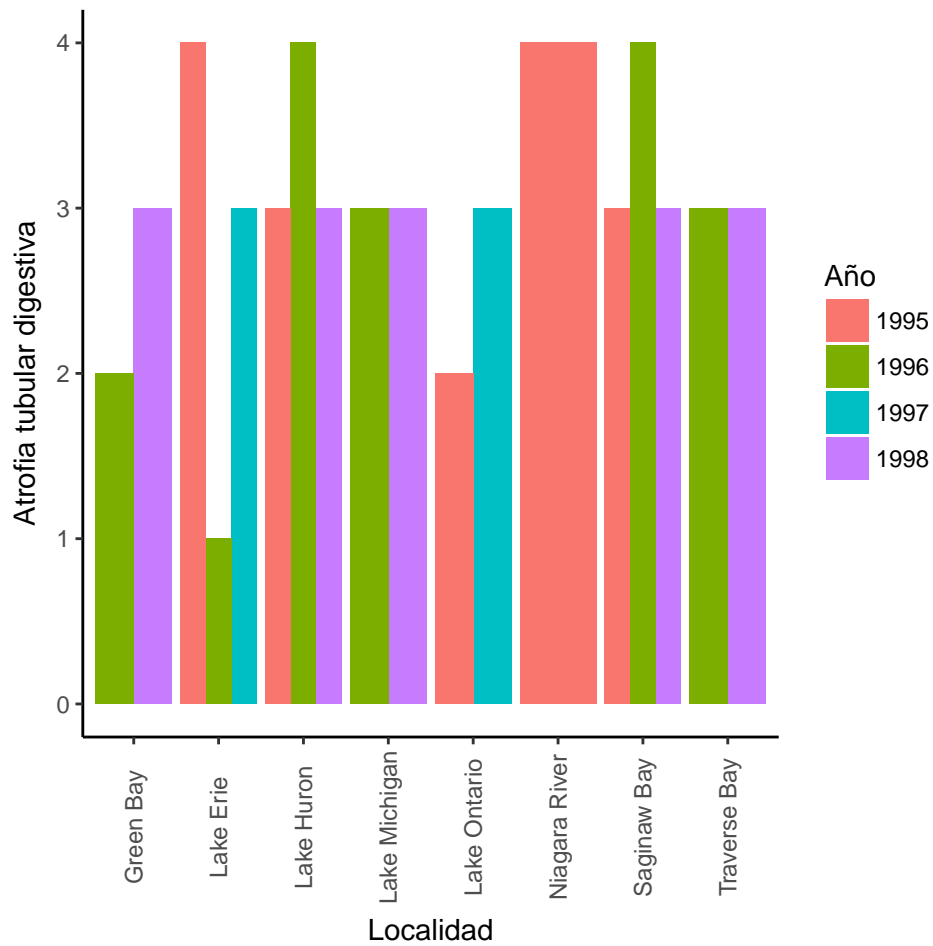


Figure 2: Figura 3. Niveles de atrofia tubular digestiva para el grupo contaminante 1 por año y localidad.

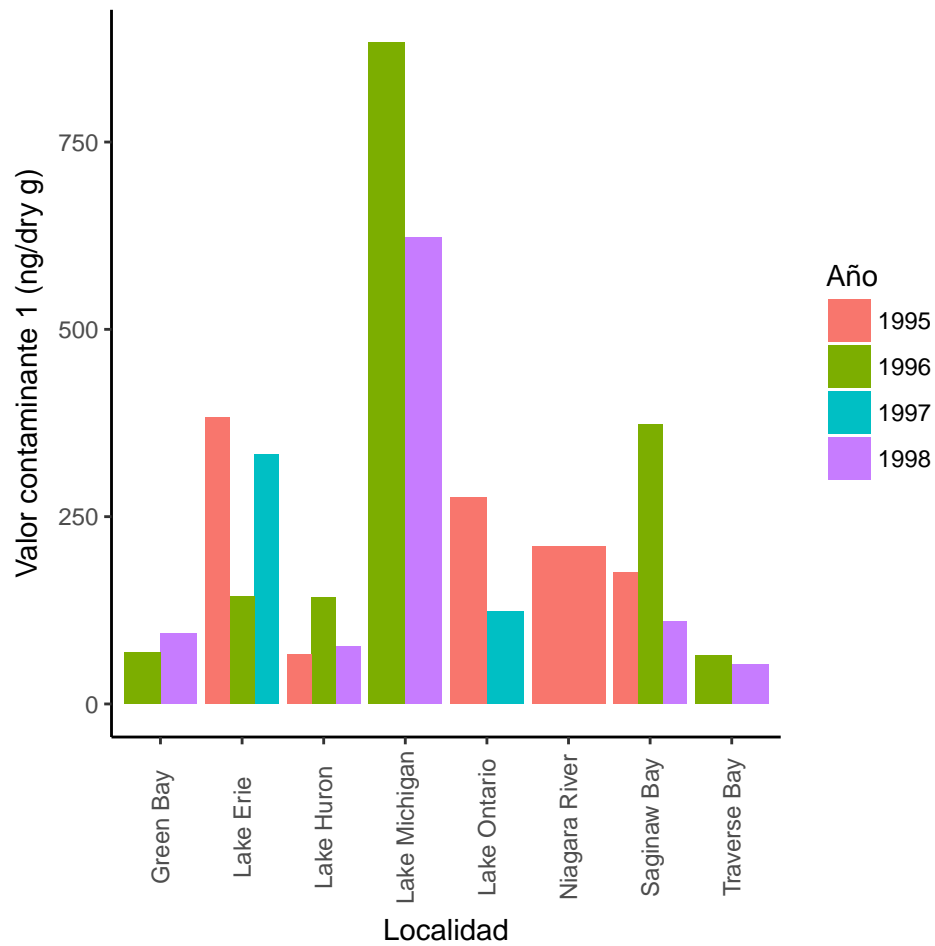


Figure 3: Figura 4. Concentración de contaminantes del grupo contaminante 1 por año y localidad.

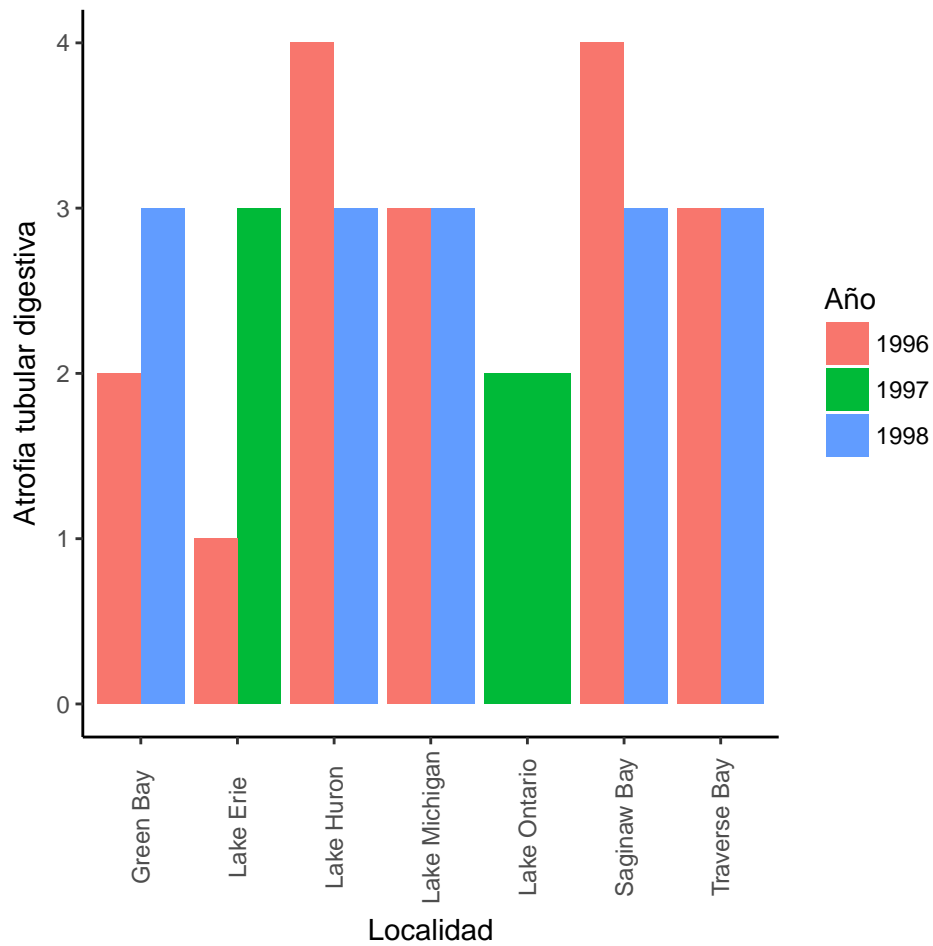


Figure 4: Figura 5. Niveles de atrofia tubular digestiva para el grupo contaminante 2 por año y localidad.

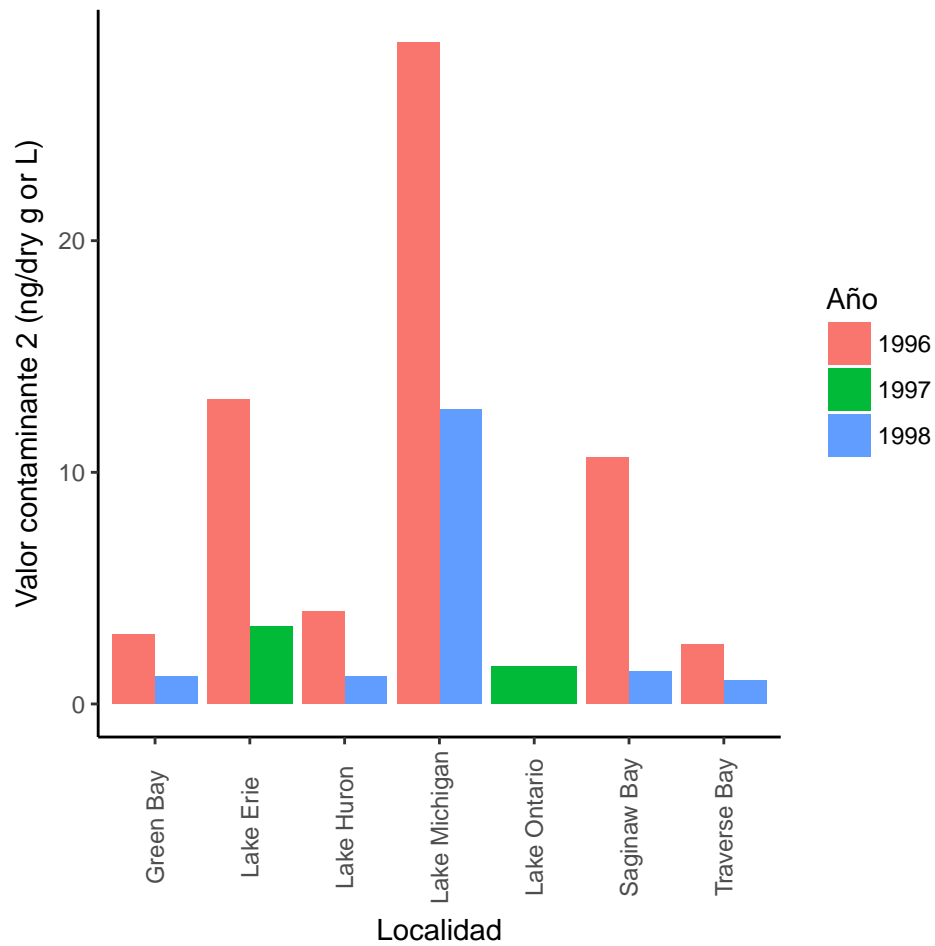


Figure 5: Figura 6. Concentración de contaminantes del grupo contaminante 2 por año y localidad.

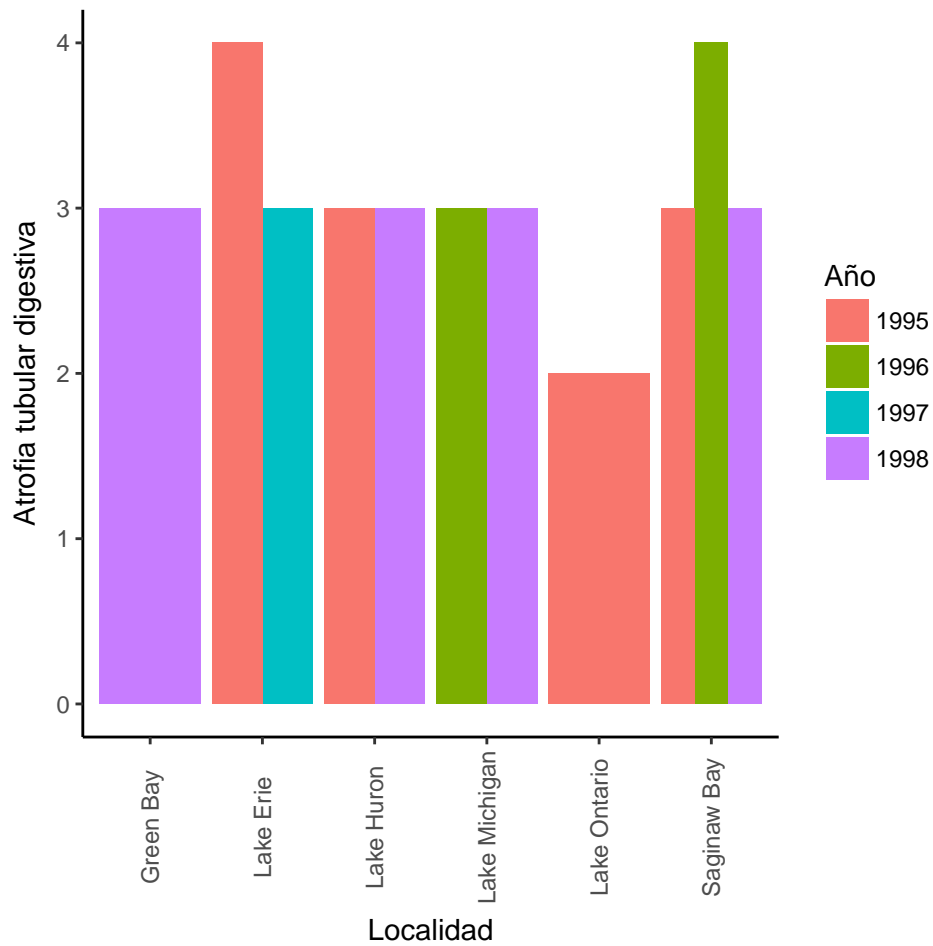


Figure 6: Figura 7. Niveles de atrofia tubular digestiva para el grupo contaminante 3 por año y localidad.



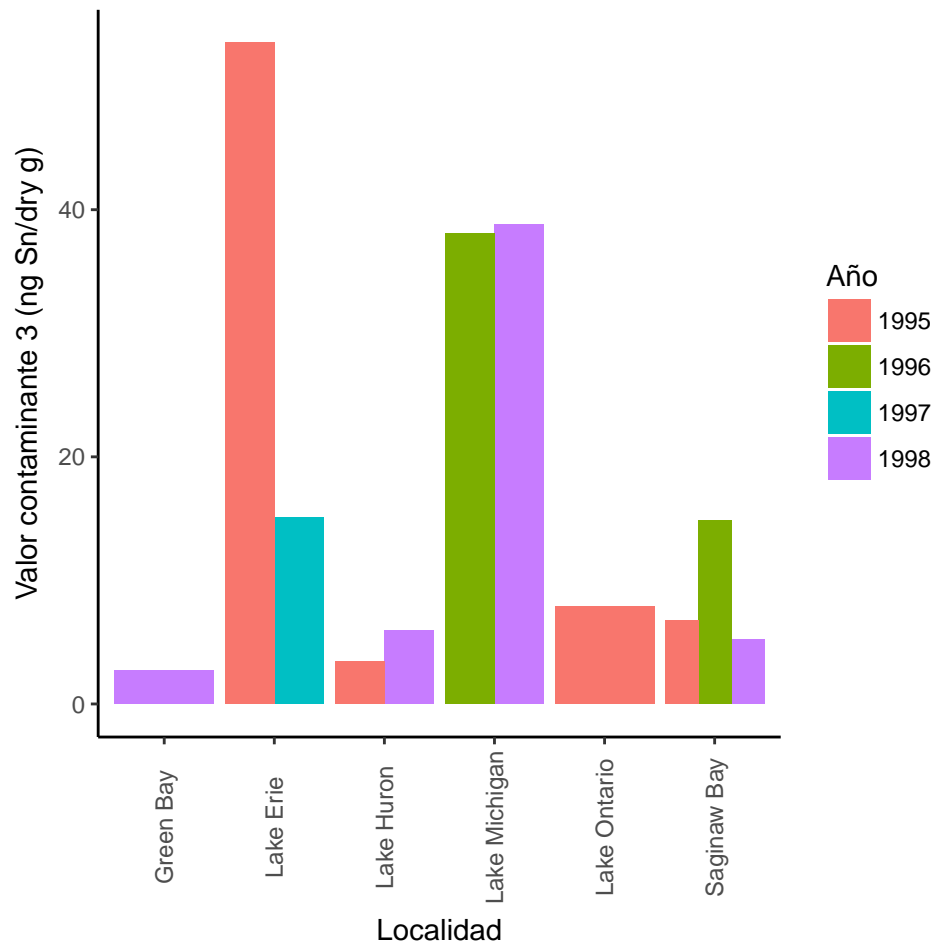


Figure 7: Figura 8. Concentración de contaminantes del grupo contaminante 3 por año y localidad.

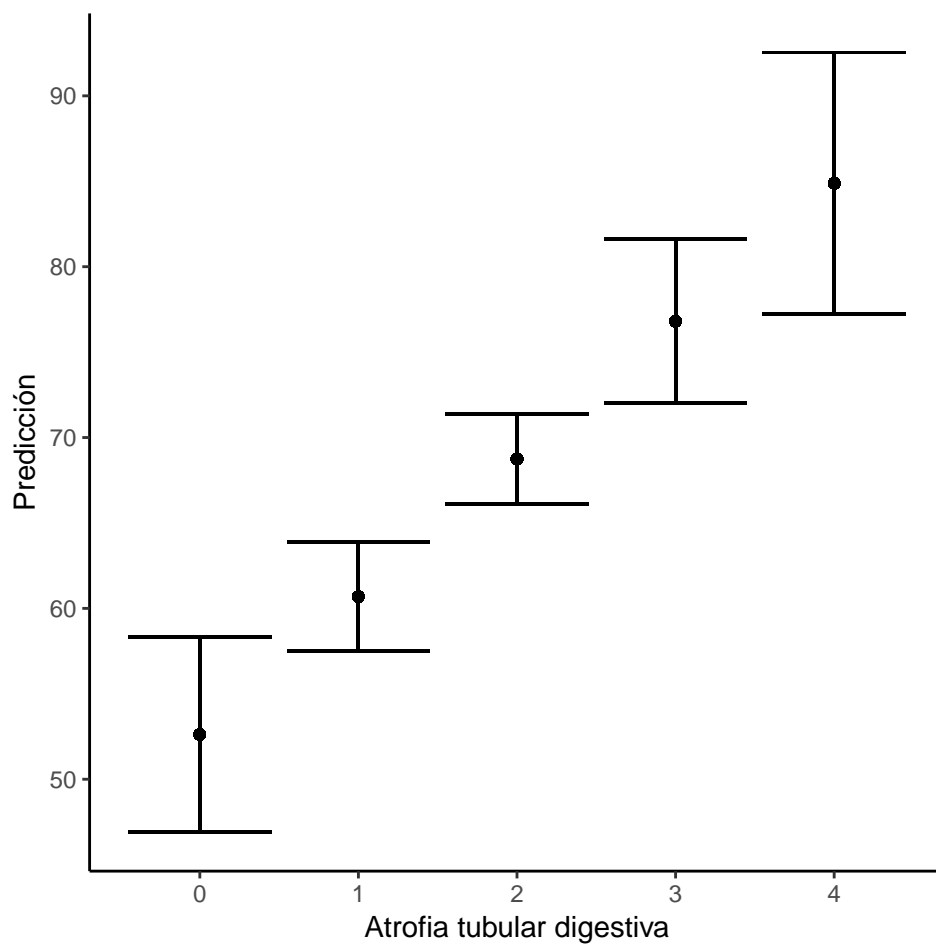


Figure 8: Figura 9. Representación gráfica del modelo con mejor ajuste para explicar las concentraciones de contaminantes.

Table 2: Tabla 2. Indices de ajuste de los modelos de medición de atrofia tubular digestiva.

r.squared	AIC	Modelo	DeltaAIC
0.1825390	44440.95	Fit24g	0.0000
0.1539941	45134.52	Fit23f	693.5757
0.1335889	45614.74	Fit22e	1173.7898
0.0990047	46393.38	Fit10	1952.4333
0.0978263	46415.72	Fit18a	1974.7690
0.0966001	46443.08	Fit15	2002.1358
0.0958863	46459.00	Fit17	2018.0498
0.0950445	46475.75	Fit8	2034.8013
0.0211337	48049.63	Fit20c	3608.6853
0.0099150	48279.24	Fit19b	3838.2968
0.0091532	48294.74	Fit9	3853.7941
0.0089019	48297.85	Fit13	3856.9039
0.0086007	48303.97	Fit14	3863.0267
0.0083744	48306.57	Fit5	3865.6248
0.0028947	48421.61	Fit21d	3980.6623
0.0013981	48449.83	Fit16	4008.8815
0.0010055	48455.75	Fit7	4014.8028
0.0004133	48467.69	Fit6	4026.7416

Table 3: Tabla 3. Indices de ajuste de los modelos de medición de los valores de contaminante.

r.squared	AIC	Modelo	DeltaAIC
0.6977957	392111.2	Fit28	0.0000
0.7126276	392523.2	Fit36	412.0803
0.6786495	393062.4	Fit25	951.2089
0.4825996	405897.3	Fit26	13786.0913
0.4703936	406439.5	Fit35	14328.3305
0.4701515	406450.6	Fit27	14339.3893
0.4698604	406464.2	Fit22	14353.0832
0.4668761	406612.6	Fit23	14501.4762
0.4661470	406651.7	Fit31	14540.5257
0.4660238	406656.3	Fit20	14545.1197
0.0040950	424337.9	Fit30	32226.6886
0.0040780	424338.3	Fit21	32227.1756
0.0043871	424343.5	Fit33	32232.3079
0.0037926	424344.5	Fit19	32233.3618
0.0041607	424348.0	Fit24	32236.8024
0.0015530	424400.7	Fit34	32289.5269
0.0012742	424406.7	Fit32	32295.5045
0.0008800	424415.9	Fit18	32304.7812
0.0002344	424434.4	Fit29	32323.2375

## Bibliografía

Hédouin, Laëtitia, Olivier Pringault, Paco Bustamante, Renaud Fichez, and Michel Warnau. 2011. “Validation of Two Tropical Marine Bivalves as Bioindicators of Mining Contamination in the New Caledonia Lagoon:

Field Transplantation Experiments.” *Water Research* 45 (2). Elsevier: 483–96.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Robinson, David, and Alex Hayes. 2018. *Broom: Convert Statistical Analysis Objects into Tidy Tibbles*. <https://CRAN.R-project.org/package=broom>.

Waykar, Bhalchandra, and Gajanan Deshmukh. 2012. “Evaluation of Bivalves as Bioindicators of Metal Pollution in Freshwater.” *Bulletin of Environmental Contamination and Toxicology* 88 (1). Springer: 48–53.

Wickham, Hadley. 2017. *Tidyverse: Easily Install and Load the 'Tidyverse'*. <https://CRAN.R-project.org/package=tidyverse>.

Wickham, Hadley, Jim Hester, and Romain Francois. 2017. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.

Xie, Yihui. 2018. *Knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.name/knitr/>.

Xie, Yihui, J.J. Allaire, and Garrett Golemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown>.