LABORATOIRE
DE SCIENCES JUDICIAIRES
ET DE MÉDECINE LÉGALE

Using R in forensic toxicology
Your automation Swiss knife

Brigitte Desharnais, Ph.D.

SOFT 2022
Monday, October 31st 2022

From raw ingredients to beautiful dishes

Data:

- Literature
- Case info
- Instruments output
- Toxicology reports

Outcome:

- Summaries to stakeholders
- Informative tools
- Actionable insight

Welcome to your data analysis kitchen!

Dann McKenzie, Flickr

Excel

SQL*LIMS v5

Welcome to your data analysis kitchen!

FREE! solution:

Programming language

https://www.r-project.org/

RStudio®

Integrated development environment (IDE)

https://www.rstudio.com/

Disclaimer

I am by no means an R absolutist.

I am a pragmatist – I use what does the job quick & easy.

… it just so happens that this is often R!

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

Graphs_Paper.R | Graphs_Paper.R

Source on Save | Run | Source

```
6  ### Upload necessary packages.
7  library(dplyr)
8  library(ggplot2)
9  library(extrafont)
10
11 ### Fig. 1
12 RED <- read.table("...", header = TRU
13
14 ggplot(data = RED, aes(x = Nb, y = urep)) +
15   geom_point(size = 1) +
16   geom_line() +
17   scale_x_continuous(name = expression("Number of measu
18
```

23:1 (Top Level) R Script

The recipe book

Environment | History | Connections

Import Dataset | 218 MiB | List

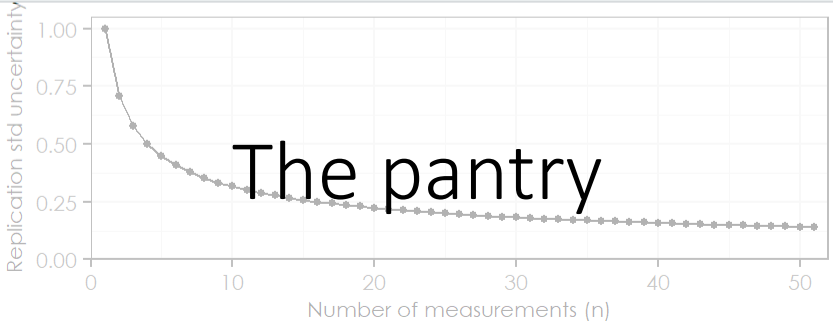
R | Global Environment

Data	
model	3 obs. of 2 variables
RED	51 obs. of 2 variables
Values	
Conc	num [1:3] 30 80 300
MU	num [1:3] 1 4 13

The countertop

Files | Plots | Packages | Help | Viewer | Presentation

Zoom | Export



The pantry

Console | Terminal | Jobs

R 3.6.1 · ~/DAC/TIAFT Bulletin/BDE - Statistics Series/8 - Advanced UM Topics/Measurement_Ur

```
+   theme_bw() +
+   theme(axis.title = element_text(size = 8, family = "Cent
ury Gothic"),
+         axis.text = element_text(size = 8, family = "Centu
ry Gothic"))
>
```

Packages: fancy appliances in your kitchen

Importation

Wrangling (structure, clean)

Visualization

Reporting

A good cook can start from any ingredient…

(Data importation)

Text files

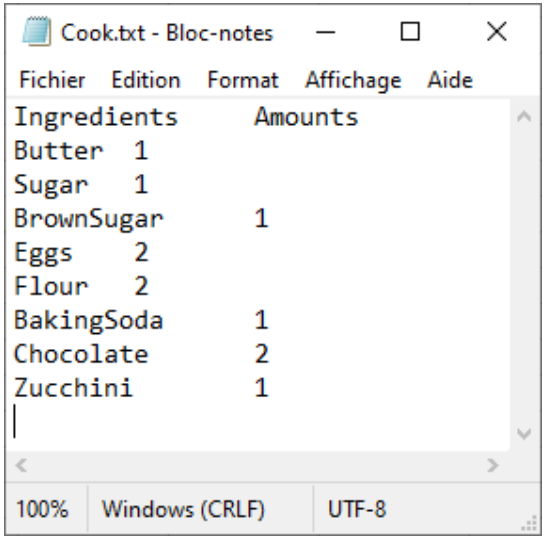
- Available in base R

```
1  Data <- read.delim("C:/Users/DESB02/RDev/Cook.txt",
2                     header = TRUE,
3                     sep = "\t")
```

Environment | **History** | **Connections**

```
Data
Data                8 obs. of 2 variables
    $ Ingredients: Factor w/ 8 levels "Bak…
    $ Amounts    : int  1 1 1 2 2 1 2 1
```

NB. Can also use `"clipboard"` instead of a file name!

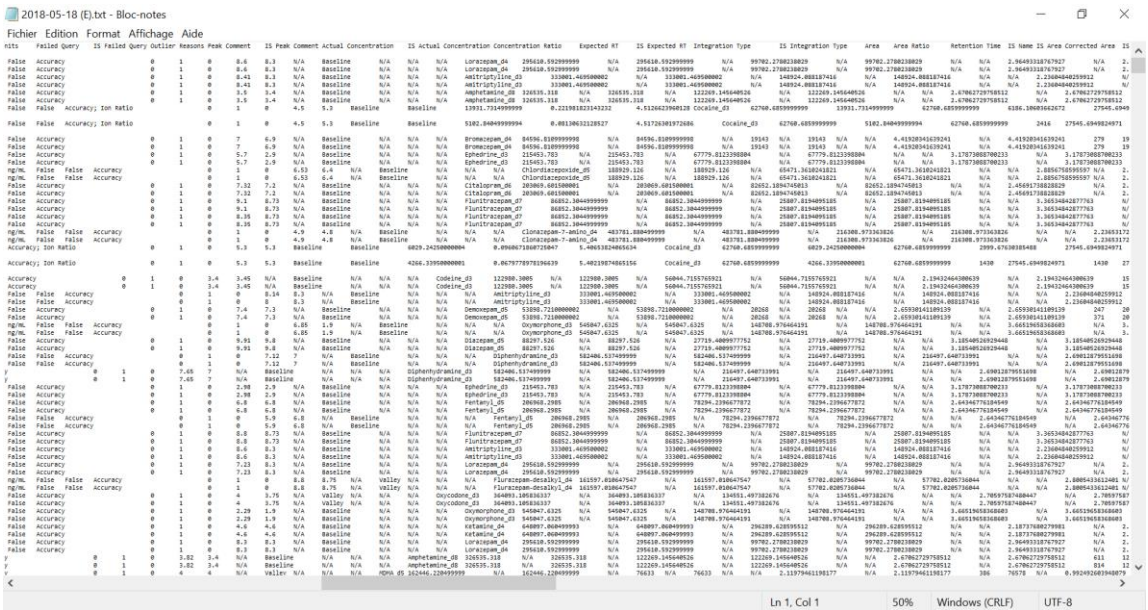


A good cook can start from any ingredient…

(Data importation)

Text files

- Of course aim is more complex files!
  - E.g., exported instrumental output (Sciex)



A good cook can start from any ingredient…

(Data importation)

Spreadsheets

- Available in base R
- Muuuuuuuch better with packages

File format	Package
.csv	readr
.xls(x)	readxl
Google Sheet	googlesheets4

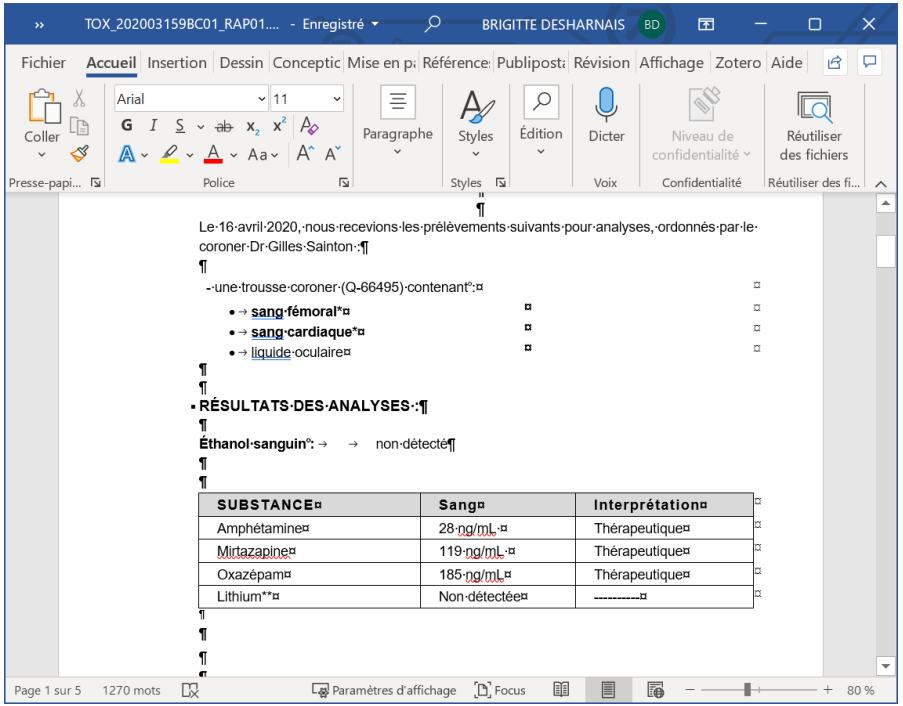
- E.g.: databases,
  exported instrumental output (Agilent)

A good cook can start from any ingredient…

(Data importation)

Text based documents

- File type: .doc(x), .pdf, .html
- Packages required
  - readtext
  - docxtractr
- E.g.: toxicology reports,
  literature,
  systems output



TOX_202003159BC01_RAP01.... - Enregistré — BRIGITTE DESHARNAIS

Fichier Accueil Insertion Dessin Conceptic Mise en p: Référence: Publipost: Révision Affichage Zotero Aide

Coller — Presse-papi... — Arial 11 — Police — Paragraphe — Styles — Édition — Dicter — Voix — Niveau de confidentialité — Confidentialité — Réutiliser des fichiers — Réutiliser des fi...

Le 16 avril 2020, nous recevions les prélèvements suivants pour analyses, ordonnés par le coroner Dr Gilles Sainton :

- une trousse coroner (Q-66495) contenant :
  - sang fémoral*
  - sang cardiaque*
  - liquide oculaire

RÉSULTATS DES ANALYSES :

Éthanol sanguin : → non détecté

SUBSTANCE	Sang	Interprétation
Amphétamine	28 ng/mL	Thérapeutique
Mirtazapine	119 ng/mL	Thérapeutique
Oxazépam	185 ng/mL	Thérapeutique
Lithium**	Non détectée	----------

Page 1 sur 5 — 1270 mots — Paramètres d'affichage — Focus — 80 %

… transform it however they want it…

(Data wrangling)

- Hands down, packages:
  - **dplyr** – select, filter, arrange, group…

Get the appliance out of the pantry

```
1  library(dplyr)
2  Data <- as_tibble(Data)
3  Dessert <- Data %>% filter(Ingredients != "Zucchini")
```

	Ingredients	Amounts
1	Butter	1
2	Sugar	1
3	BrownSugar	1
4	Eggs	2
5	Flour	2
6	BakingSoda	1
7	Chocolate	2
8	Zucchini	1

	Ingredients	Amounts
1	Butter	1
2	Sugar	1
3	BrownSugar	1
4	Eggs	2
5	Flour	2
6	BakingSoda	1
7	Chocolate	2

… transform it however they want it…

(Data wrangling)

- Hands down, packages:
  - **dplyr** – select, filter, arrange, group…
  - **tidyr** – gather, spread, separate, unite…

```
1  library(tidyr)
2  Cleaned <- Results %>% pivot_wider(names_from = Matrix,
3                                     values_from = Concentration)
```

	File	Matrix	Analyte	Concentration
1	2022-A	CB	Amphetamine	20
2	2022-A	FB	Amphetamine	19
3	2022-A	UR	Amphetamine	39
4	2022-B	CB	Amphetamine	32
5	2022-B	VH	Amphetamine	24
6	2022-B	GC	Amphetamine	5
7	2022-B	UR	Amphetamine	36
8	2022-C	CB	Ethanol	22
9	2022-C	FB	Ethanol	29

	File	Analyte	CB	FB	UR	VH	GC
1	2022-A	Amphetamine	20	19	39	*NA*	*NA*
2	2022-B	Amphetamine	32	*NA*	36	24	5
3	2022-C	Ethanol	22	29	*NA*	*NA*	*NA*

(And you can use dplyr to filter – on the same line if you want!)

… transform it however they want it…

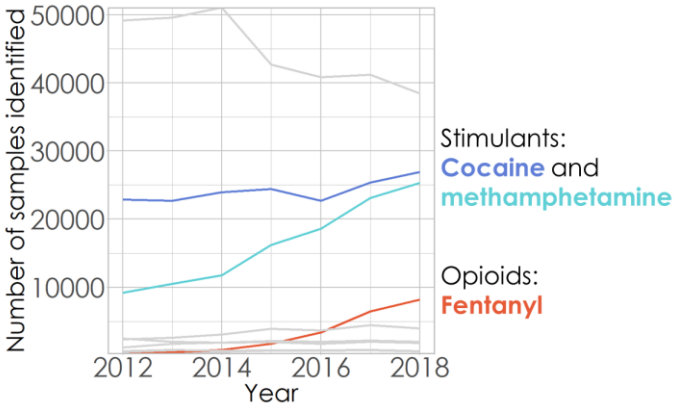
(Data wrangling)

- Hands down, packages:
  - **dplyr** – select, filter, arrange, group…
  - **tidyr** – gather, spread, separate, unite…
  - stringr – search & modify character strings (text)
  - lubridate – work with time and dates
  - janitor – cleaning names, empty column/lines

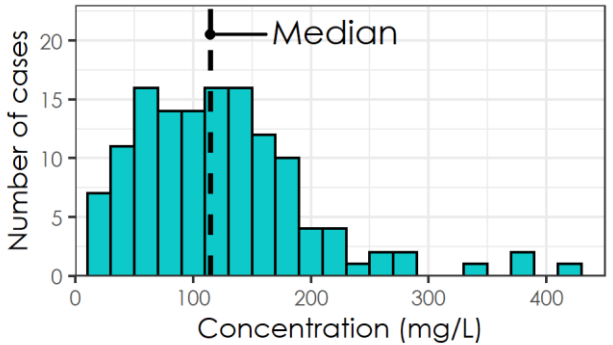
… to create beautiful and delicious plates!

(Visualization & reporting)

- For graphs, ggplot2 & associated packages:
  - extrafont – use alternative fonts for labelling
  - export – send & manipulate in e.g. Power Point (!!!)
  - gganimate – animated plots
  - plotly – graphs that can be manipulated (zoom, save, etc.)



TIAFT Bulletin (2019) 49 (4), 14-26

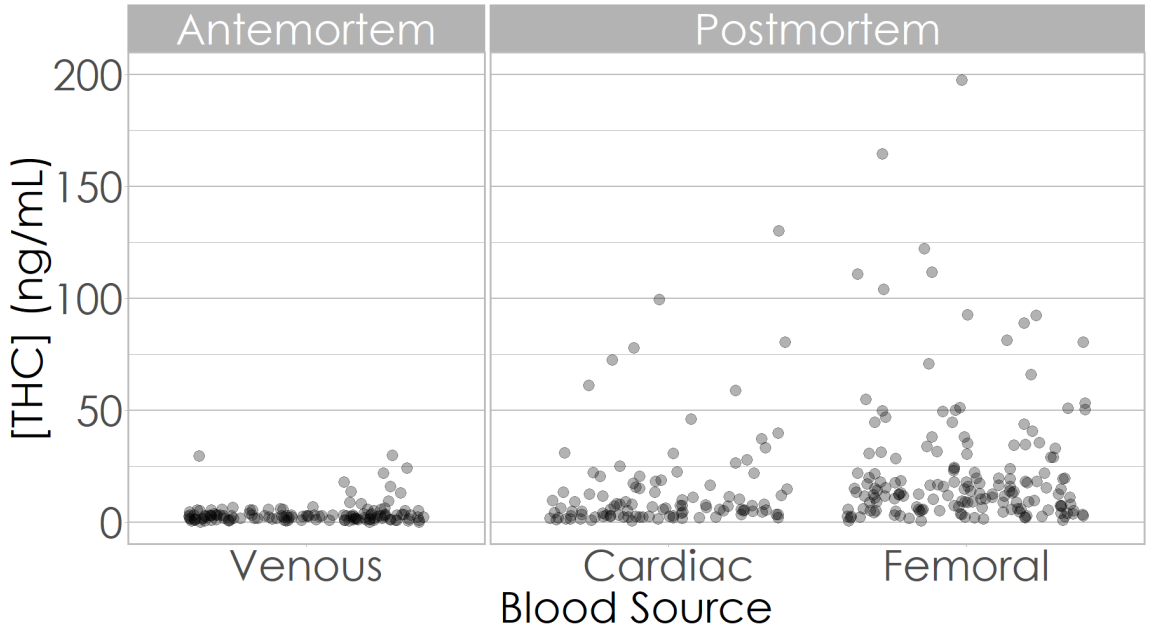


TIAFT Bulletin (2020) 50 (4), 11-19

… to create beautiful and delicious plates!

(Visualization & reporting)

- Faceted plots

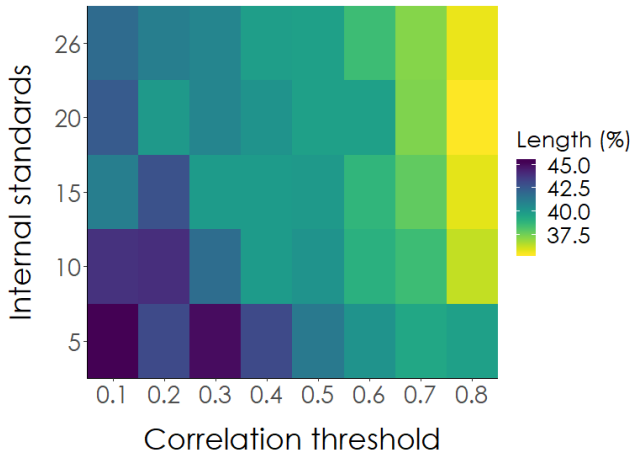


TIAFT Bulletin (2019) 49 (4), 14-26

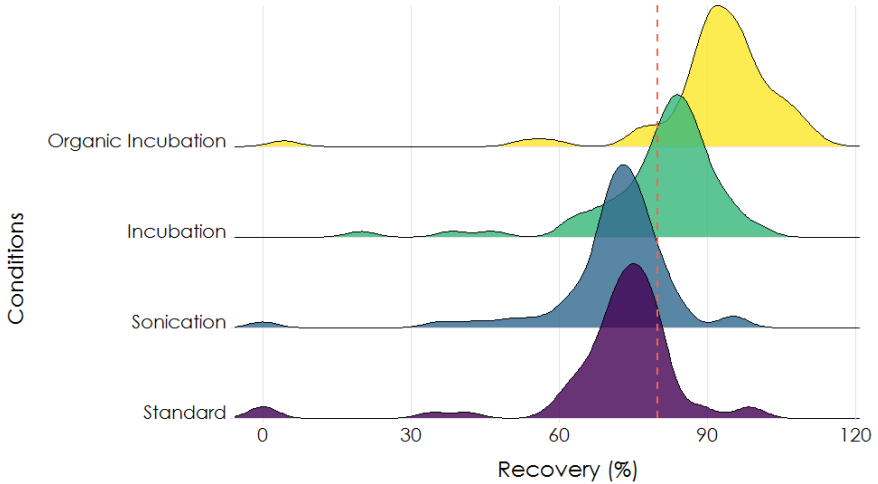
… to create beautiful and delicious plates!

(Visualization & reporting)

- Heat maps
- Density
- Ridges



TIAFT 2019 Annual Meeting

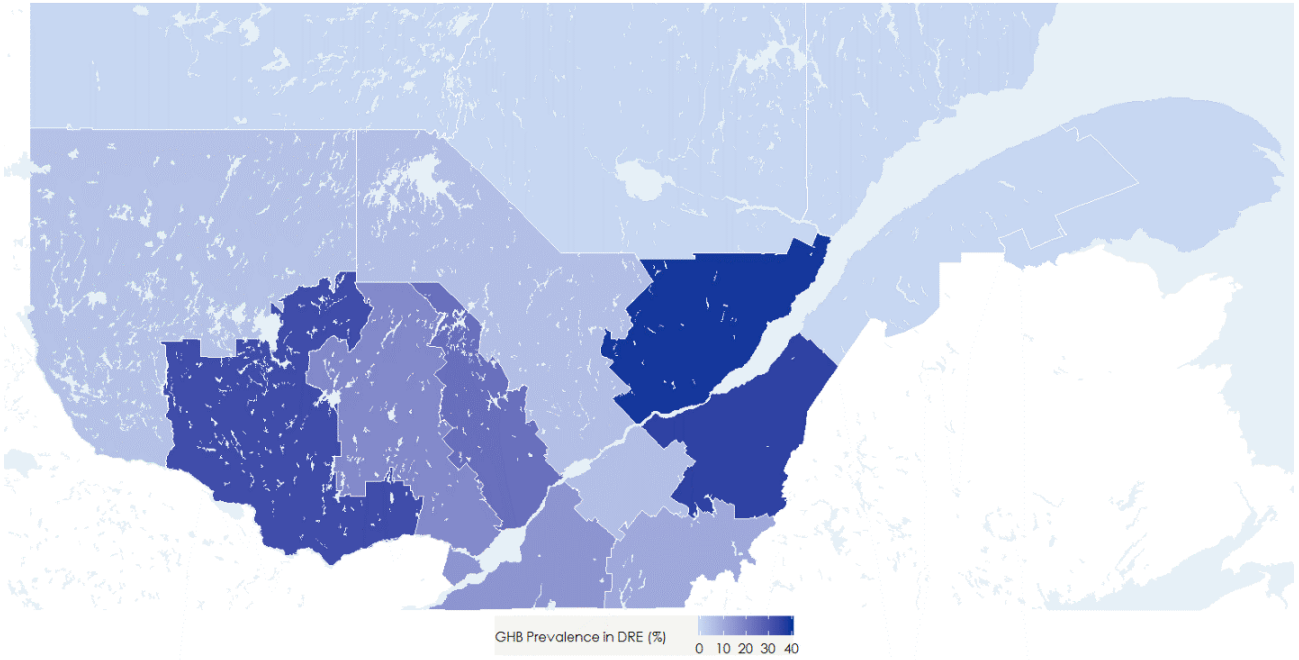


Forensic Sci Int (2020) 317, 110506

… to create beautiful and delicious plates!

(Visualization & reporting)

- Maps

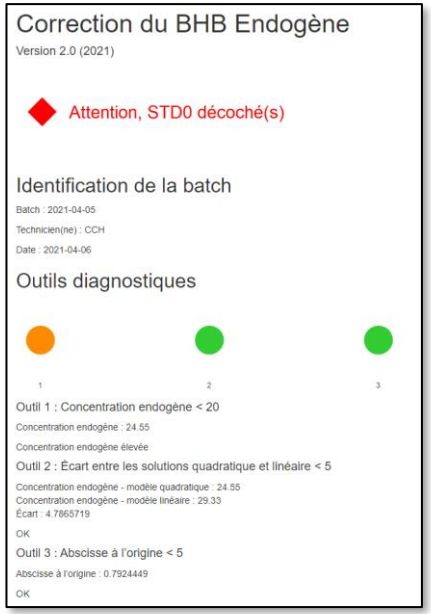


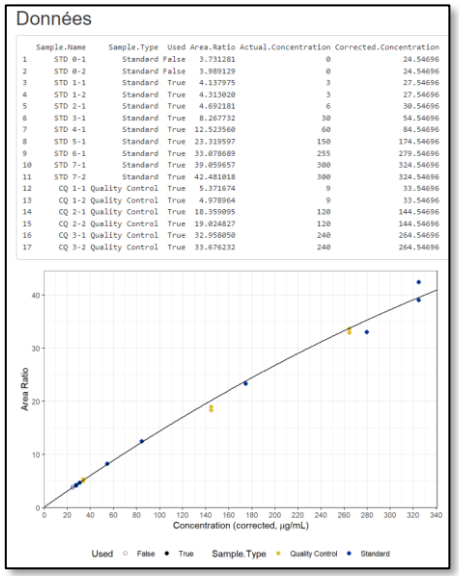
SOFT 2019 Annual Meeting

… to create beautiful and delicious plates!

(Visualization & reporting)

- For standardized reports:
  - rmarkdown – generate .html or .pdf reports
  - quarto – new generation rmarkdown, includes WYSIWYG editor





… to create beautiful and delicious plates!

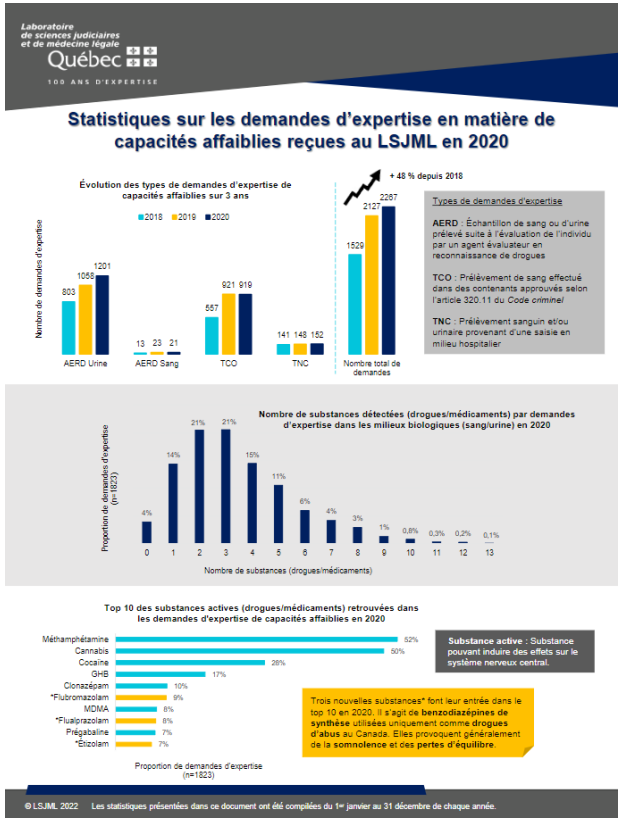
(Visualization & reporting)

- For interactive reports, dashboards:
  - shiny – the foundation
  - shinydashboard
  - bs4dash
  - flexdashboard
  - …
- Interactive elements can be integrated in reports!
- Examples at https://shiny.rstudio.com/gallery/

A five course meal

Applications of R in our laboratory

Hors d'oeuvre: annual summary reports



Appetizer: population distributions

GHB - Étude de population

AUTHOR
Brigitte Desharnais, Laboratoire de sciences judiciaires et de médecine légale

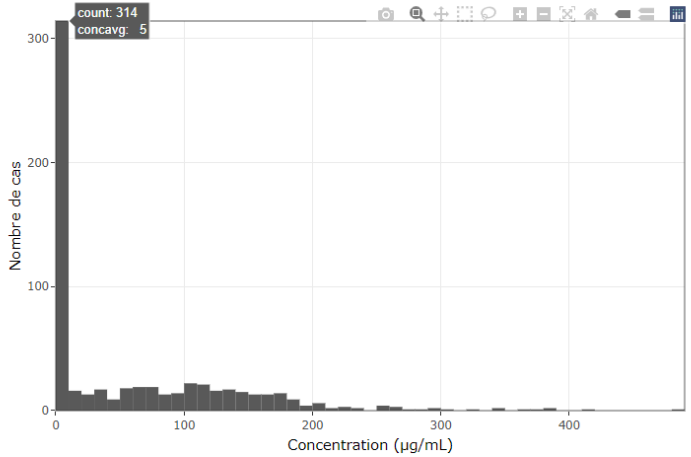
Données entre le 2017-01-22 et le 2022-07-22 (date de batch).

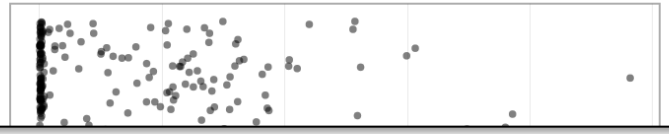
Répartition des données

Milieu	Ante mortem	Post mortem
Sang veineux (SV)	630	66
Sang fémoral (SF)	6	9788
Sang cardiaque (SC)	0	2379
Plasma/Sérum (PS)	81	21
Urine (UR)	715	1314
Liquide occulaire (LO)	0	458

Ante mortem

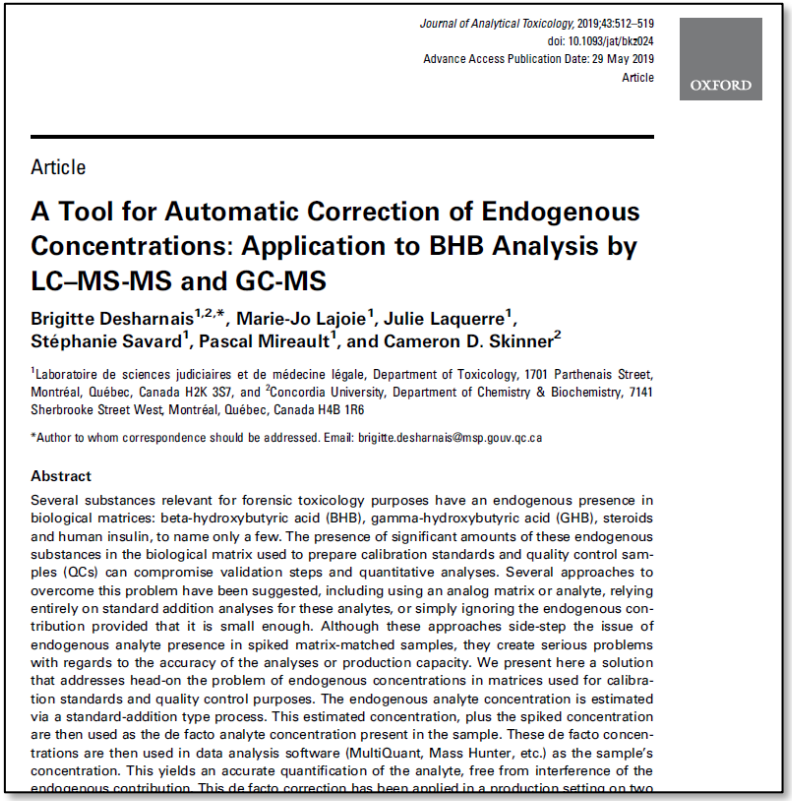
Sang veineux





Salad: correction of endogenous content in spiking matrix

- Endogenous analytes create a validation challenge with regards to spiking matrix
- Possible solutions include
  - Standard addition
  - Surrogate matrix
  - Surrogate analyte
  - Ignoring the problem (!)



Journal of Analytical Toxicology, 2019;43:512–519
doi: 10.1093/jat/bkz024
Advance Access Publication Date: 29 May 2019
Article

OXFORD

Article

A Tool for Automatic Correction of Endogenous Concentrations: Application to BHB Analysis by LC–MS-MS and GC-MS

Brigitte Desharnais[1,2,*], Marie-Jo Lajoie[1], Julie Laquerre[1], Stéphanie Savard[1], Pascal Mireault[1], and Cameron D. Skinner[2]

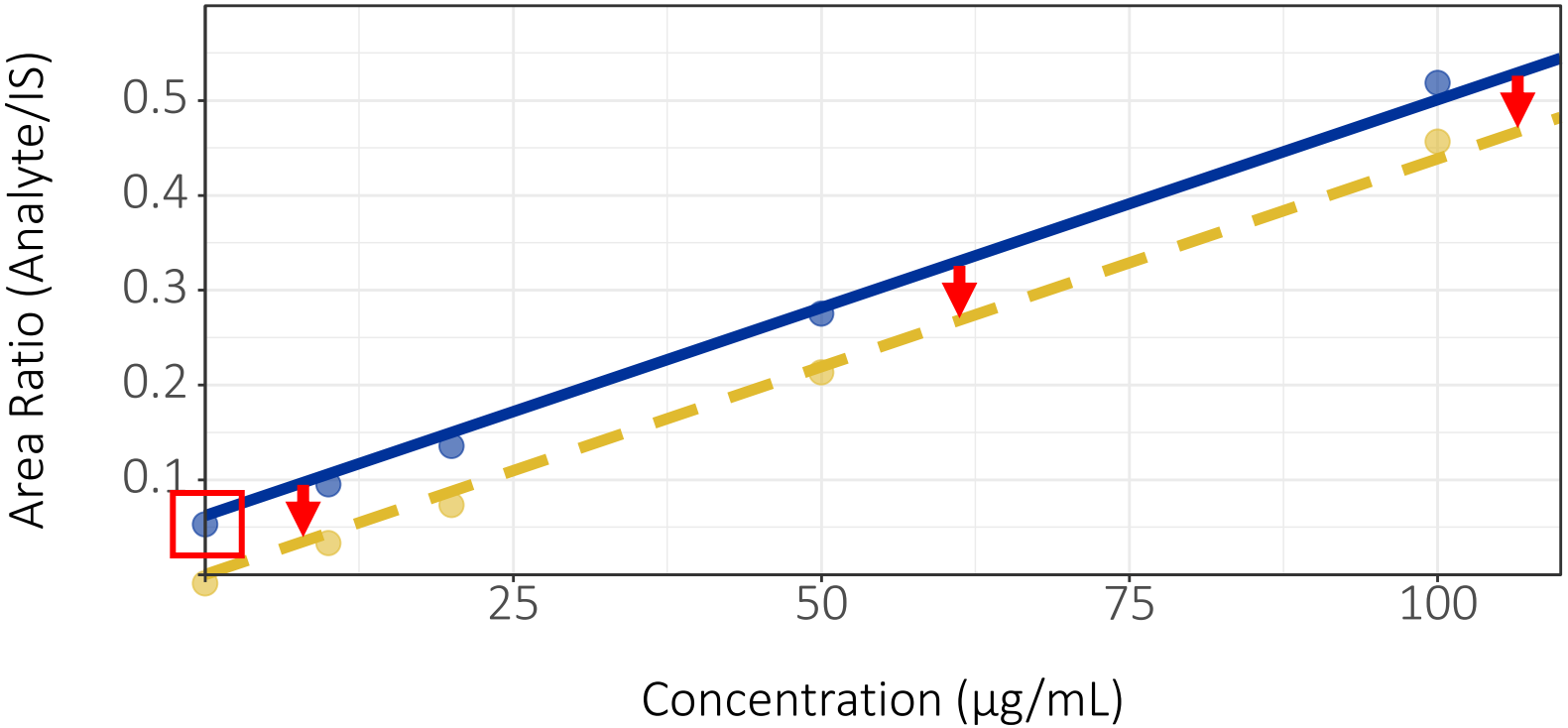
[1]Laboratoire de sciences judiciaires et de médecine légale, Department of Toxicology, 1701 Parthenais Street, Montréal, Québec, Canada H2K 3S7, and [2]Concordia University, Department of Chemistry & Biochemistry, 7141 Sherbrooke Street West, Montréal, Québec, Canada H4B 1R6

*Author to whom correspondence should be addressed. Email: brigitte.desharnais@msp.gouv.qc.ca

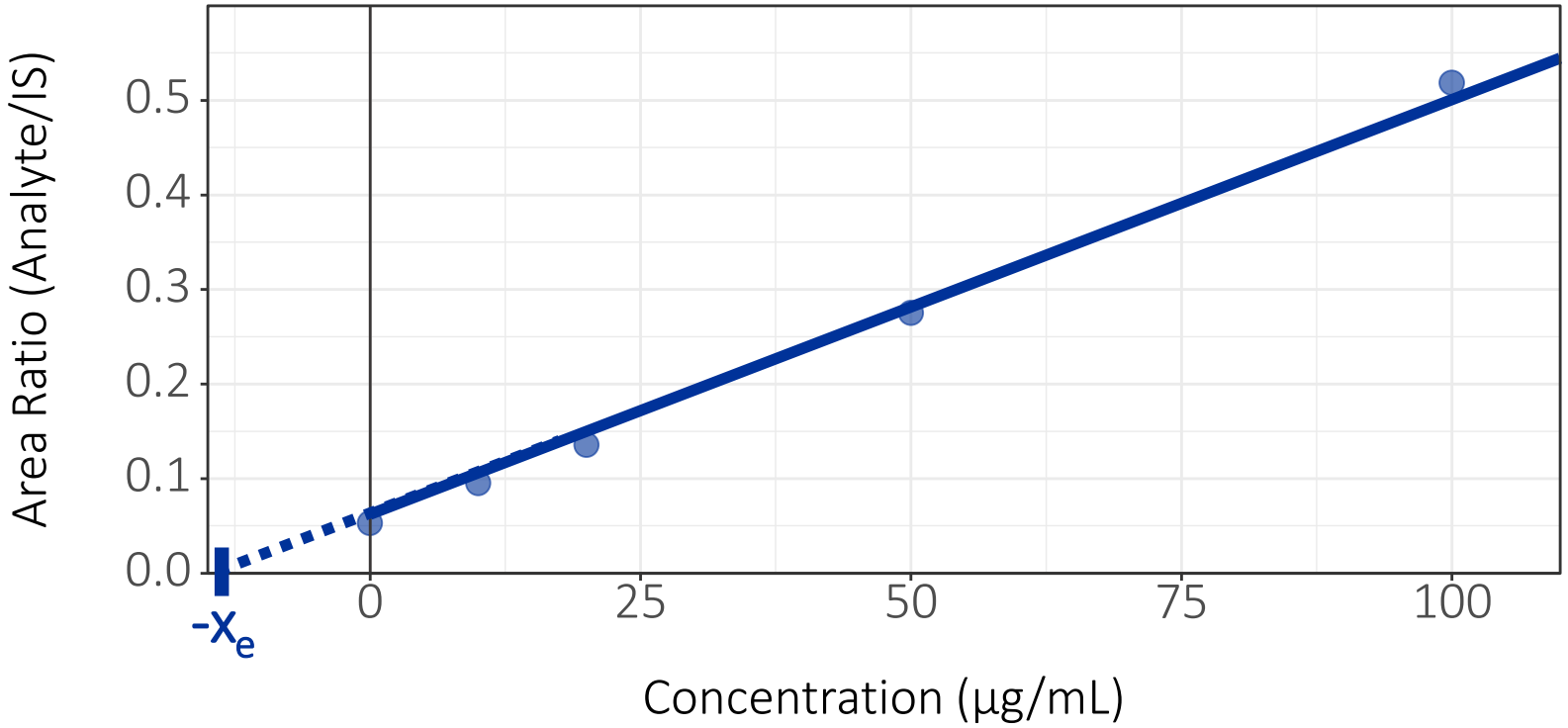
Abstract

Several substances relevant for forensic toxicology purposes have an endogenous presence in biological matrices: beta-hydroxybutyric acid (BHB), gamma-hydroxybutyric acid (GHB), steroids and human insulin, to name only a few. The presence of significant amounts of these endogenous substances in the biological matrix used to prepare calibration standards and quality control samples (QCs) can compromise validation steps and quantitative analyses. Several approaches to overcome this problem have been suggested, including using an analog matrix or analyte, relying entirely on standard addition analyses for these analytes, or simply ignoring the endogenous contribution provided that it is small enough. Although these approaches side-step the issue of endogenous analyte presence in spiked matrix-matched samples, they create serious problems with regards to the accuracy of the analyses or production capacity. We present here a solution that addresses head-on the problem of endogenous concentrations in matrices used for calibration standards and quality control purposes. The endogenous analyte concentration is estimated via a standard-addition type process. This estimated concentration, plus the spiked concentration are then used as the de facto analyte concentration present in the sample. These de facto concentrations are then used in data analysis software (MultiQuant, Mass Hunter, etc.) as the sample's concentration. This yields an accurate quantification of the analyte, free from interference of the endogenous contribution. This de facto correction has been applied in a production setting on two

Salad: correction of endogenous content in spiking matrix

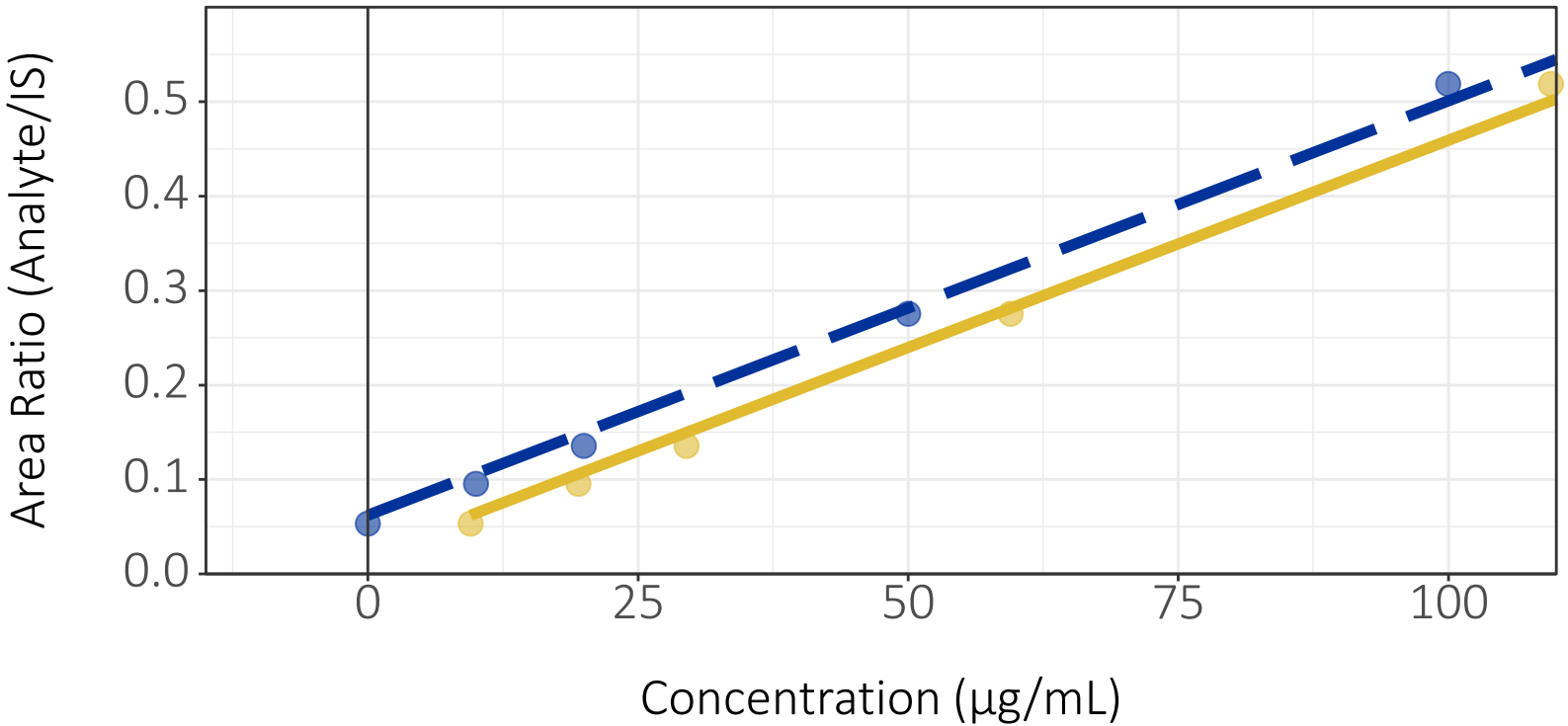


Salad: correction of endogenous content in spiking matrix



Correction: $x_r = x_s + x_e$

Salad: correction of endogenous content in spiking matrix



Correction: $x_r = x_s + x_e$

MultiQuant - [[MQ4] Results Table (2018-04-27 LV PPT3 Sg.qsession)]

File Edit Process Window Help

CYP (D)\2017-07 (D)

Components & Groups IS

Pregabaline d6
Acide salicylique d4
Butalbital d5
GHB d6
Phenobarbital d5
Phenytoine d10
Topiramate d12

All Analytes
Acetaminophene 1
Acetaminophene 2
Carbamazepine 1
Carbamazepine 2
Celecoxib 1
Celecoxib 2
Gabapentine 1
Gabapentine 2
Ibuprofene 1
Lamotrigine 1
Lamotrigine 2
Levetiracetam 1
Levetiracetam 2
Methylphenidate 1
Methylphenidate 2
Naproxene 1
Naproxene 2
Pregabaline 1
Pregabaline 2
Carbamazepine-10.11-dihydroxy 1
Carbamazepine-10.11-dihydroxy 2
Celecoxib-COOH 1
Celecoxib-COOH 2
Ibuprofene-COOH 1
Ibuprofene-COOH 2
Oxcarbazepine 1
Oxcarbazepine 2
Phenytoine-p-hydroxy 1
Phenytoine-p-hydroxy 2
Pregabaline-N-methyl 1
Pregabaline-N-methyl 2
Psilocine 1
Psilocine 2
Acide salicylique 1
Acide salicylique 2
Acide valproique 1
Barbital 1
Barbital 2
BHB 1
BHB 2
Butalbital 1
Butalbital 2
GHB 1
GHB 2
Phenobarbital 1
Phenobarbital 2
Phenytoine 1
Phenytoine 2
Topiramate 1
Topiramate 2
Acetaminophene-glucuronide 1
Acetaminophene-glucuronide 2
Acide valproique-glucuronide 1
Acide valproique-glucuronide 2
Naproxene-O-desmethyl 1
Naproxene-O-desmethyl 2
Propofol-glucuronide 1
Propofol-glucuronide 2

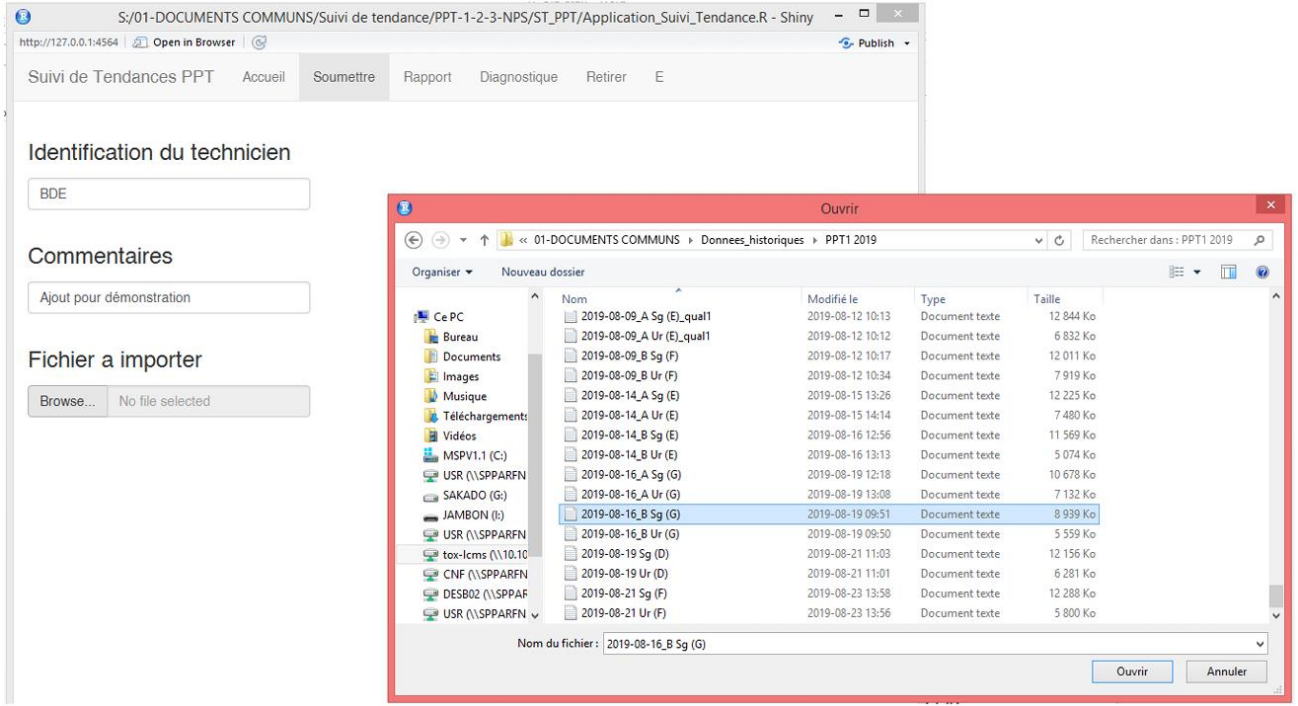
Standards & QCs

Index	Sample Name	Sample Type	IS	Component Name	IS Name	Actual Concentrati_	Area	IS Area	Area Ratio	Retention Time	IS Reten_ Time	Width at 50%	Used	Calculated Concentration	Accuracy
486	STD 0-1	Standard		BHB 1	GHB_d6	0.000	1.769e4	3.081e5	0.0574	0.98	0.80	0.04		0.199	N/A
559	STD 0-2	Standard		BHB 1	GHB_d6	0.000	1.649e4	3.001e5	0.0549	1.00	0.81	0.03		0.164	N/A
632	STD 1-1	Standard		BHB 1	GHB_d6	3.000	7.901e4	2.971e5	0.2659	0.99	0.80	0.04	✓	3.124	104.14
705	STD 1-2	Standard		BHB 1	GHB_d6	3.000	7.413e4	2.899e5	0.2557	0.99	0.80	0.04	✓	2.981	99.36
778	STD 2-1	Standard		BHB 1	GHB_d6	6.000	1.256e5	2.865e5	0.4386	0.99	0.81	0.04	✓	5.546	92.43
851	STD 3-1	Standard		BHB 1	GHB_d6	30.000	6.798e5	3.081e5	2.2063	0.99	0.80	0.04	✓	30.280	100.93
924	STD 4-1	Standard		BHB 1	GHB_d6	60.000	1.291e6	2.859e5	4.5144	0.99	0.81	0.04	✓	62.433	104.06
997	STD 5-1	Standard		BHB 1	GHB_d6	150.000	3.154e6	2.936e5	10.7398	0.99	0.81	0.04	✓	148.353	98.90
1070	STD 6-1	Standard		BHB 1	GHB_d6	255.000	5.336e6	2.799e5	19.0620	0.99	0.80	0.04	✓	261.456	102.53
1143	STD 7-1	Standard		BHB 1	GHB_d6	300.000	6.075e6	2.776e5	21.8830	0.99	0.80	0.04	✓	299.354	99.78
1216	STD 7-2	Standard		BHB 1	GHB_d6	300.000	5.600e6	2.611e5	21.4504	0.99	0.81	0.04	✓	293.556	97.85
1289	CQ 1-1	Quality Co..		BHB 1	GHB_d6	9.000	1.894e5	2.868e5	0.6604	0.99	0.81	0.03	✓	8.654	96.16
1362	CQ 1-2	Quality Co..		BHB 1	GHB_d6	9.000	1.785e5	2.795e5	0.6384	1.00	0.81	0.04	✓	8.347	92.74
1435	CQ 2-1	Quality Co..		BHB 1	GHB_d6	120.000	2.810e6	2.948e5	9.5323	0.99	0.80	0.04	✓	131.777	109.81
1508	CQ 2-2	Quality Co..		BHB 1	GHB_d6	120.000	2.862e6	2.960e5	9.6700	0.99	0.81	0.04	✓	133.670	111.39
1581	CQ 3-1	Quality Co..		BHB 1	GHB_d6	240.000	5.450e6	3.014e5	18.0816	0.99	0.80	0.04	✓	248.233	103.43
1654	CQ 3-2	Quality Co..		BHB 1	GHB_d6	240.000	5.113e6	2.734e5	18.7021	0.99	0.81	0.04	✓	256.604	106.92

FR 5:25 PM 5/3/2018

Main course: trend monitoring

LC-MS/MS method, 3 injections, 125 analytes, 7 STDs, 3x3 QCs…



S:/01-DOCUMENTS COMMUNS/Suivi de tendance/PPT-1-2-3-NPS/ST_PPT/Application_Suivi_Tendance.R - Shiny

http://127.0.0.1:5769 | Open in Browser | Publish

Suivi de Tendances PPT | Accueil | Soumettre | Rapport | Diagnostique | Retirer | E

Plage de dates

2019-10-11 to 2019-11-11

Instrument(s)

- [x] D
- [x] E
- [x] F
- [x] G

Variable Etudiee

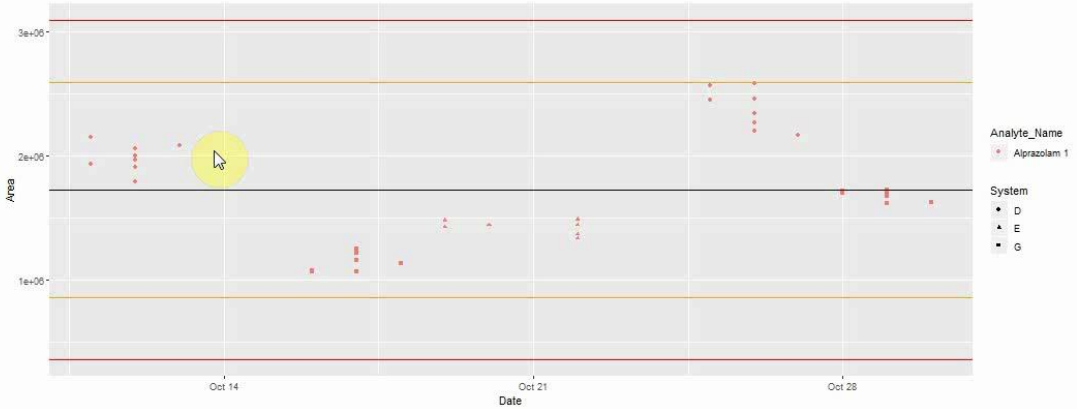
Area

Type Echantillon

REF

Analyte(s)

- [x] Alprazolam
- [ ] Amitriptyline
- [ ] Amphetamine
- [ ] Benzoylecgonine
- [ ] Bromazepam
- [ ] Bupropion
- [ ] Chlordiazepoxide
- [ ] Citalopram
- [ ] Clobazam
- [ ] Clonazepam
- [ ] Clonazepam-7-amino
- [ ] Cocaine
- [ ] Codeine



Batch	Date	Methode	Commentaires
398	2019-10-16	PPT1	std 3 exclu pour plusieurs analytes (amphet, bupropion, clonaz-7a, methamph, sertra, venla-O-desm
403	2019-10-16	PPT3	topiramate qual
407	2019-10-19	PPT1	nouvelle sln SS3 et ds critères. Olanzapine LLQ 10 ng/mL
416	2019-10-22	PPT1	olanzapine pas de STD 1-1 et 1-2. Norpseudo %CV REF à 32%
419	2019-10-21	PPT2	1 seul CO pour THC
421	2019-10-22	PPT3	manque de sensibilité pour psilocine dans 2 REF
422	2019-10-22	PPT3	manque de sensibilité pour psilocine dans 2 REF
423	2019-10-17	PPT1	NC19-283 sln expirée sans impact. STD 3-1 exclus de tous les analytes. Amphet, bromaz et oxaz 2x moins sensibles.
424	2019-10-17	PPT1	NC19-283 sln expirée sans impact.
425	2019-10-17	PPT2	NC19-283 sln expirée sans impact. STD 3-1 exclus de tous les analytes.
428	2019-10-28	PPT1	olanzapine STD1-2 exclu
430	2019-10-29	PPT2	%CV Aire THC-d3 57%
433	2019-10-28	PPT3	Topiramate pas STD-1 sensibilité
441	2019-10-25	PPT2	THC pas STD 1
442	2019-10-25	PPT2	pas de THC CO-urine-1
450	2019-10-29	PPT1	contient 3ng/mL citalopram

Main course: trend monitoring

Periodic PDF report for accreditation requirements

Laboratoire de sciences judiciaires et de médecine légale Québec

Rapport de suivi des tendances pour les methodes PPT 1, 2, 3 et NPS (MA-08-TOX-16, 17 et 18)

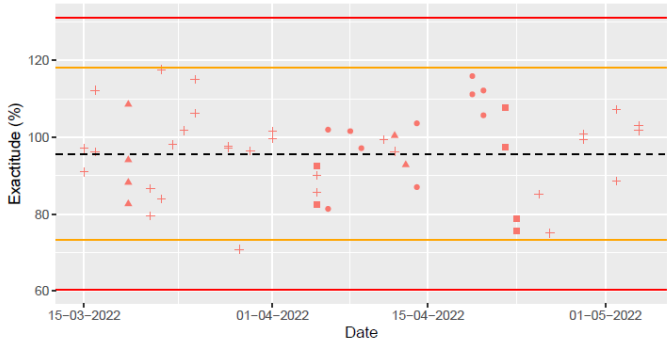
Identification

Rapport genere le 2022-05-11 par CDO.
Couvre la periode du 2022-03-14 au 2022-05-08.

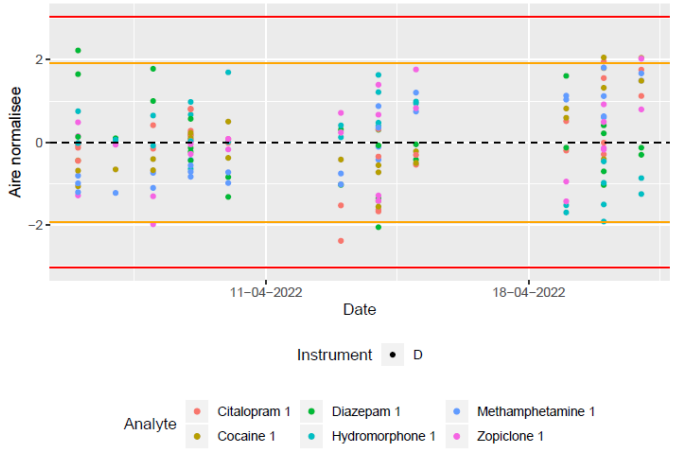
Ce rapport couvre 16.5 sequences uniques de PPT1, 15.5 sequences uniques de PPT2, 13 sequences uniques de PPT3 et 30 sequences uniques de PPT-NPS.

PPT 1

Exactitude Methamphetamine (CQ 1)



Sensibilite PPT 1 LC-MS/MS D (REF)

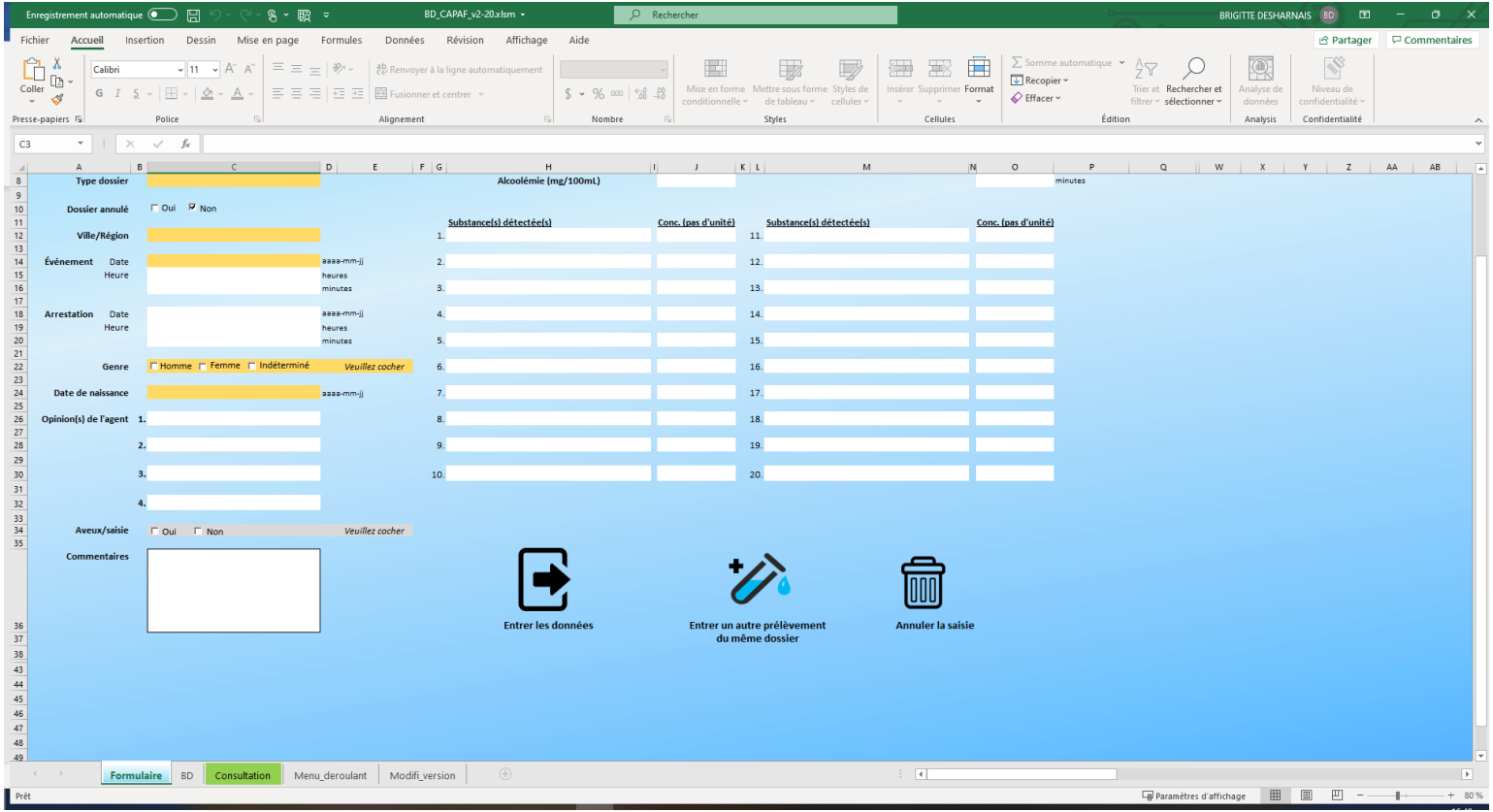


Critere 1 ("warning lines"). Le nombre de points sous la LWL 3, pour un maximum permis de 9. Ce critere est donc respecte: TRUE.

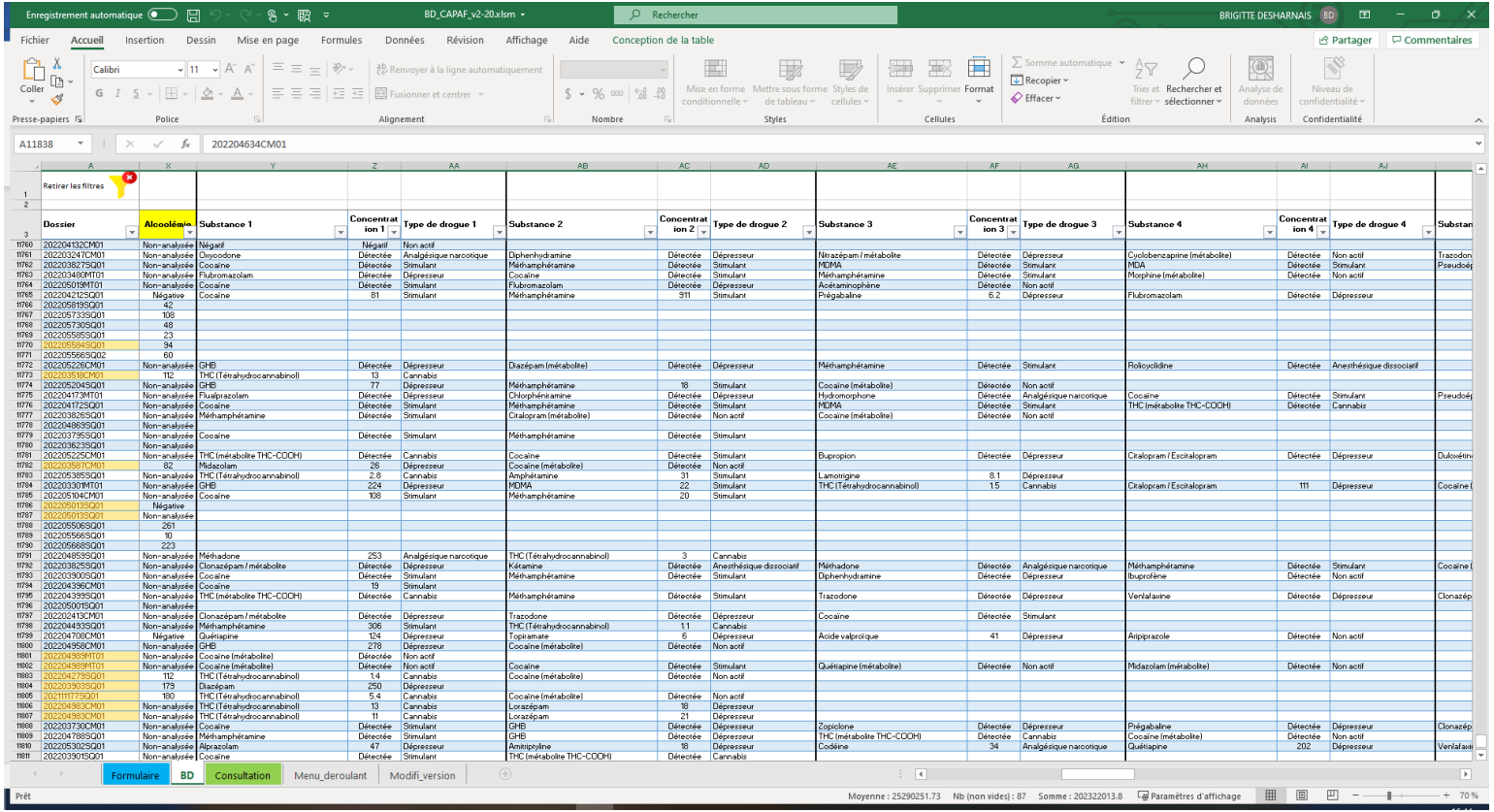
Critere 2 ("action lines"). Le nombre de points sous la LAL est de 0, pour un maximum permis de 1. Ce critere est donc respecte: TRUE.

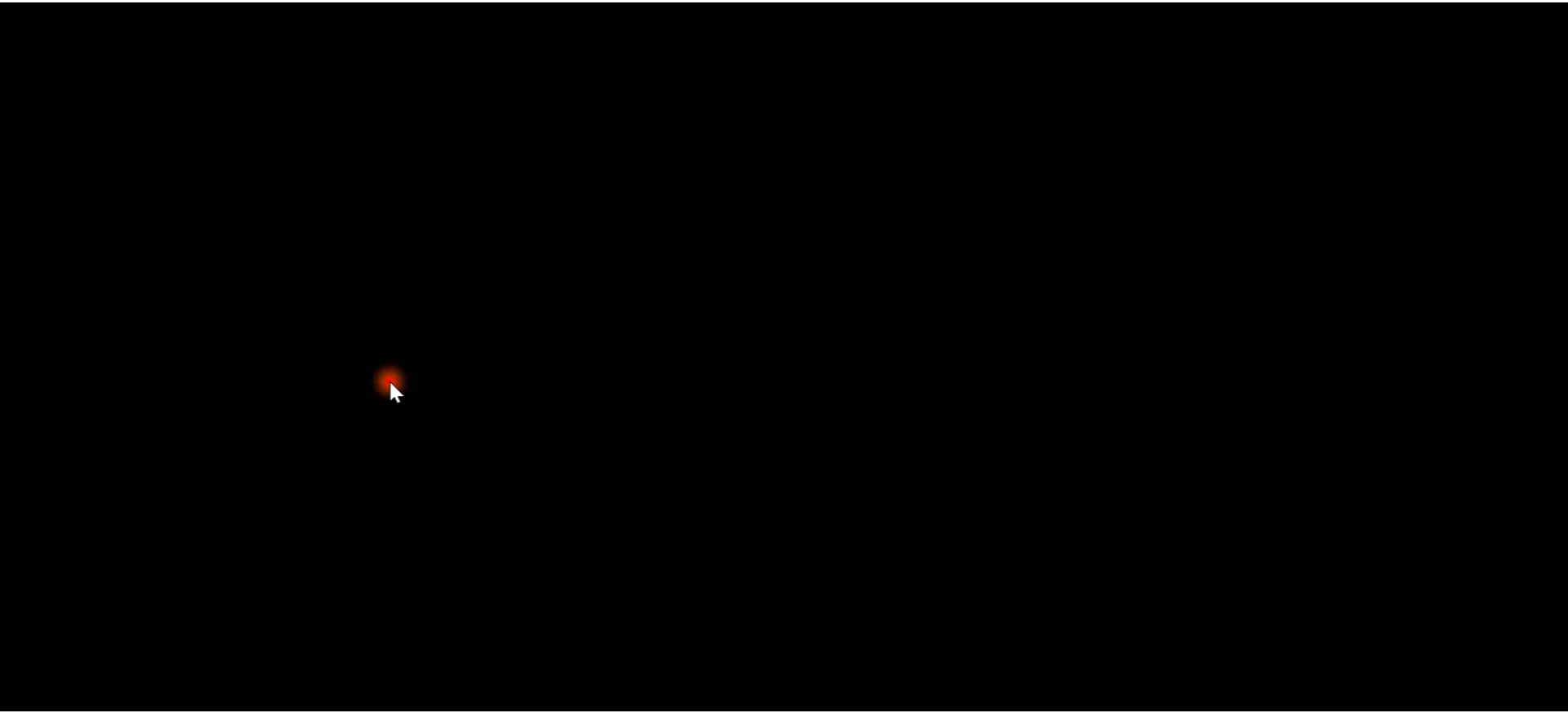
Critere 3 (pas de diminution significative). La regression lineaire effectuee sur les donnees a une pente de 0.0393239, et un intervalle de confiance allant de -0.0098989 a 0.0885467. Puisque l'intervalle de confiance n'inclut pas que des valeurs negatives, ce critere est respecte: TRUE.

Dessert: DUID interactive dashboard



Dessert: DUID interactive dashboard





A teaser of our upcoming menus

- Just a sample of what we are doing with R!
- **Take home message: if you have an idea, there's a way to do it in R.**
- But we want to explore more…
- Docking with other tools
  - Excel with VBA
  - Power BI (Microsoft)
- Version control using GitHub (would help with accreditation requirements!)

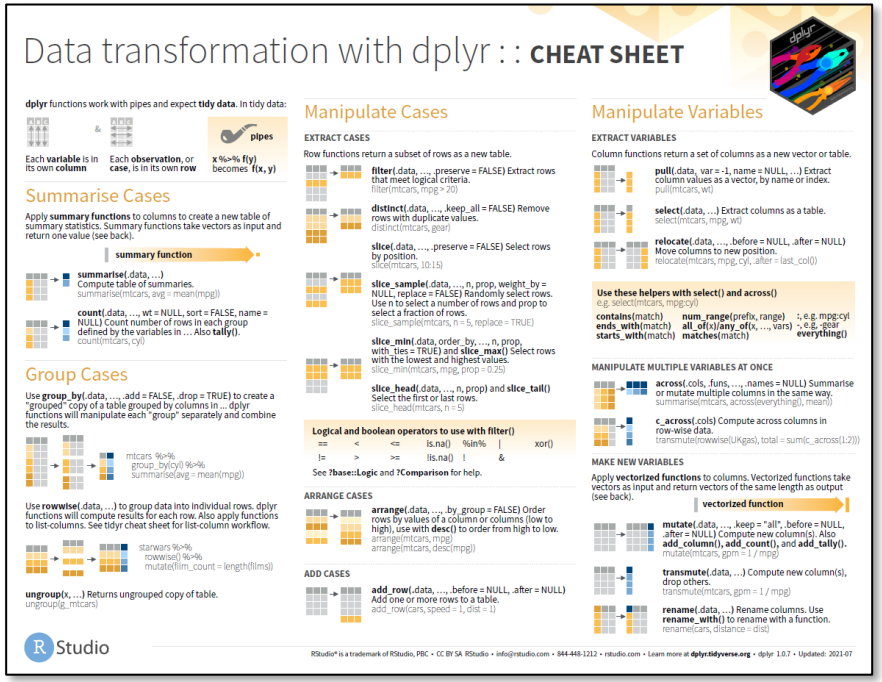
Want to get cooking?

Do you need to learn it early to be good with it?

Want to get cooking? Several free resources online!

R Studio Cheat Sheets – super dense, but it's all there…

> https://www.rstudio.com/resources/cheatsheets/



Data transformation with dplyr : : **CHEAT SHEET**

dplyr functions work with pipes and expect **tidy data**. In tidy data:

Each **variable** is in its own **column** & Each **observation**, or **case**, is in its own **row**

pipes x %>% f(y) becomes f(x, y)

Summarise Cases

Apply **summary functions** to columns to create a new table of summary statistics. Summary functions take vectors as input and return one value (see back).

summary function

summarise(.data, …) Compute table of summaries.
summarise(mtcars, avg = mean(mpg))

count(.data, …, wt = NULL, sort = FALSE, name = NULL) Count number of rows in each group defined by the variables in … Also **tally()**.
count(mtcars, cyl)

Group Cases

Use **group_by**(.data, …, .add = FALSE, .drop = TRUE) to create a "grouped" copy of a table grouped by columns in … dplyr functions will manipulate each "group" separately and combine the results.

```
mtcars %>%
  group_by(cyl) %>%
  summarise(avg = mean(mpg))
```

Use **rowwise**(.data, …) to group data into individual rows. dplyr functions will compute results for each row. Also apply functions to list-columns. See tidyr cheat sheet for list-column workflow.

```
starwars %>%
  rowwise() %>%
  mutate(film_count = length(films))
```

ungroup(x, …) Returns ungrouped copy of table.
ungroup(g_mtcars)

Manipulate Cases

EXTRACT CASES

Row functions return a subset of rows as a new table.

filter(.data, …, .preserve = FALSE) Extract rows that meet logical criteria.
filter(mtcars, mpg > 20)

distinct(.data, …, .keep_all = FALSE) Remove rows with duplicate values.
distinct(mtcars, gear)

slice(.data, …, .preserve = FALSE) Select rows by position.
slice(mtcars, 10:15)

slice_sample(.data, …, n, prop, weight_by = NULL, replace = FALSE) Randomly select rows. Use n to select a number of rows and prop to select a fraction of rows.
slice_sample(mtcars, n = 5, replace = TRUE)

slice_min(.data, order_by, …, n, prop, with_ties = TRUE) and **slice_max()** Select rows with the lowest and highest values.
slice_min(mtcars, mpg, prop = 0.25)

slice_head(.data, …, n, prop) and **slice_tail()** Select the first or last rows.
slice_head(mtcars, n = 5)

Logical and boolean operators to use with filter()

==	<	<=	is.na()	%in%	|	xor()
!=	>	>=	!is.na()	!	&	

See **?base::Logic** and **?Comparison** for help.

ARRANGE CASES

arrange(.data, …, .by_group = FALSE) Order rows by values of a column or columns (low to high), use with **desc()** to order from high to low.
arrange(mtcars, mpg)
arrange(mtcars, desc(mpg))

ADD CASES

add_row(.data, …, .before = NULL, .after = NULL) Add one or more rows to a table.
add_row(cars, speed = 1, dist = 1)

Manipulate Variables

EXTRACT VARIABLES

Column functions return a set of columns as a new vector or table.

pull(.data, var = -1, name = NULL, …) Extract column values as a vector, by name or index.
pull(mtcars, wt)

select(.data, …) Extract columns as a table.
select(mtcars, mpg, wt)

relocate(.data, …, .before = NULL, .after = NULL) Move columns to new position.
relocate(mtcars, mpg, cyl, .after = last_col())

Use these helpers with select() and across()
e.g. select(mtcars, mpg:cyl)

contains(match) **num_range**(prefix, range) **:**, e.g. mpg:cyl
ends_with(match) **all_of**(x)/**any_of**(x, …, vars) **-**, e.g. -gear
starts_with(match) **matches**(match) **everything()**

MANIPULATE MULTIPLE VARIABLES AT ONCE

across(.cols, .funs, …, .names = NULL) Summarise or mutate multiple columns in the same way.
summarise(mtcars, across(everything(), mean))

c_across(.cols) Compute across columns in row-wise data.
transmute(rowwise(UKgas), total = sum(c_across(1:2)))

MAKE NEW VARIABLES

Apply **vectorized functions** to columns. Vectorized functions take vectors as input and return vectors of the same length as output (see back).

vectorized function

mutate(.data, …, .keep = "all", .before = NULL, .after = NULL) Compute new column(s). Also **add_column()**, **add_count()**, and **add_tally()**.
mutate(mtcars, gpm = 1 / mpg)

transmute(.data, …) Compute new column(s), drop others.
transmute(mtcars, gpm = 1 / mpg)

rename(.data, …) Rename columns. Use **rename_with()** to rename with a function.
rename(cars, distance = dist)

RStudio

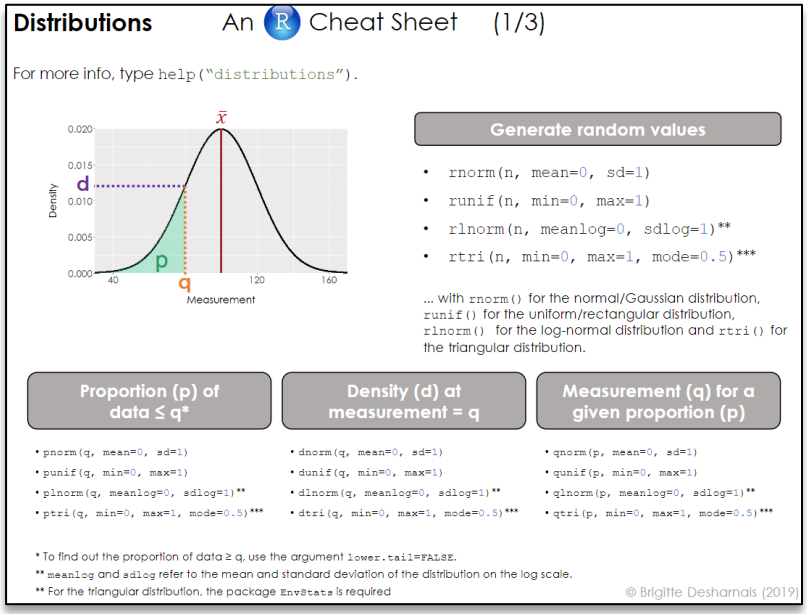
RStudio® is a trademark of RStudio, PBC • CC BY SA RStudio • info@rstudio.com • 844-448-1212 • rstudio.com • Learn more at **dplyr.tidyverse.org** • dplyr 1.0.7 • Updated: 2021-07

Want to get cooking? Several free resources online!

TIAFT Bulletin Statistics Series, including R Cheat Sheets – less dense

(but more stats oriented)

> https://1drv.ms/u/s!At1r-QnPAVZ0jjVGlQA3ex84qiRl?e=UJKKj5



Distributions An R Cheat Sheet (1/3)

For more info, type `help("distributions")`.

Generate random values

- `rnorm(n, mean=0, sd=1)`
- `runif(n, min=0, max=1)`
- `rlnorm(n, meanlog=0, sdlog=1)**`
- `rtri(n, min=0, max=1, mode=0.5)***`

... with `rnorm()` for the normal/Gaussian distribution, `runif()` for the uniform/rectangular distribution, `rlnorm()` for the log-normal distribution and `rtri()` for the triangular distribution.

Proportion (p) of data ≤ q*	Density (d) at measurement = q	Measurement (q) for a given proportion (p)
• `pnorm(q, mean=0, sd=1)`	• `dnorm(q, mean=0, sd=1)`	• `qnorm(p, mean=0, sd=1)`
• `punif(q, min=0, max=1)`	• `dunif(q, min=0, max=1)`	• `qunif(p, min=0, max=1)`
• `plnorm(q, meanlog=0, sdlog=1)**`	• `dlnorm(q, meanlog=0, sdlog=1)**`	• `qlnorm(p, meanlog=0, sdlog=1)**`
• `ptri(q, min=0, max=1, mode=0.5)***`	• `dtri(q, min=0, max=1, mode=0.5)***`	• `qtri(p, min=0, max=1, mode=0.5)***`

* To find out the proportion of data ≥ q, use the argument `lower.tail=FALSE`.

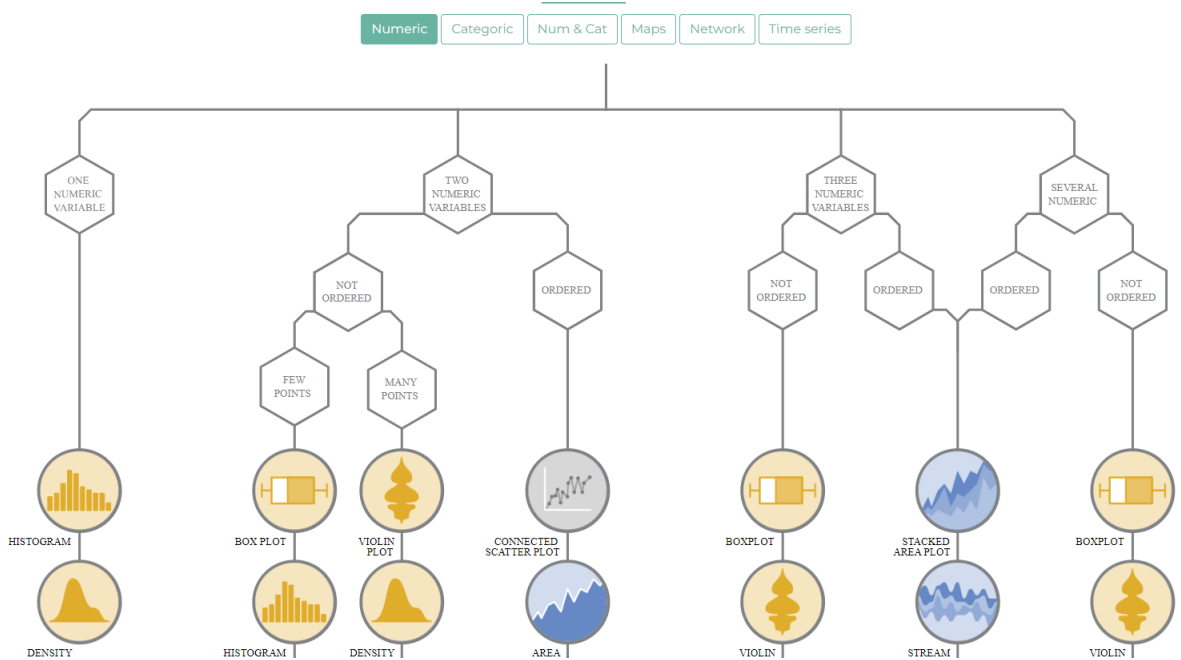
** `meanlog` and `sdlog` refer to the mean and standard deviation of the distribution on the log scale.

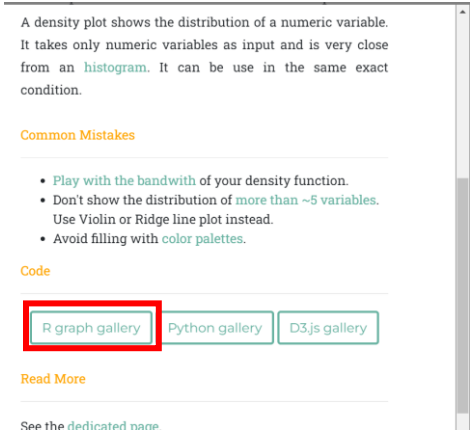
*** For the triangular distribution, the package `EnvStats` is required

Want to get cooking? Several free resources online!

From Data to Viz

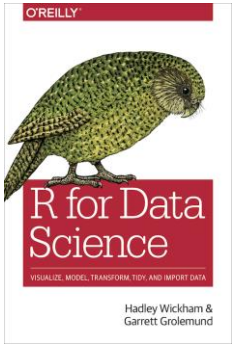
> data-to-viz.com





Want to get cooking? Several free resources online!

R Bootcamp – gentle introduction by Oregon Health & Science University
> https://r-bootcamp.netlify.app/



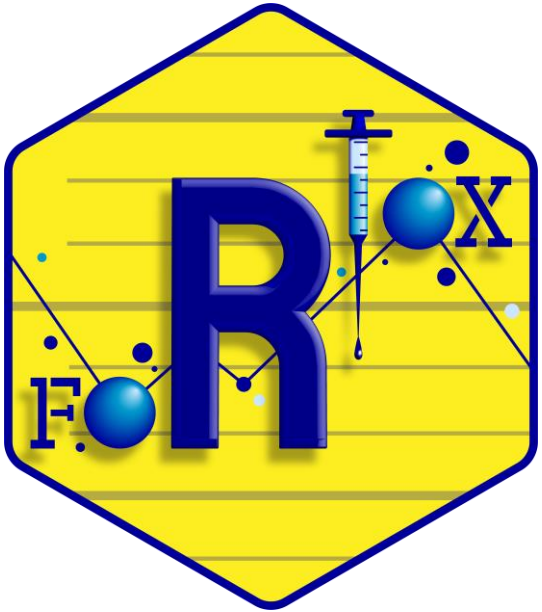
R for Data Science (H. Wickham & G. Grolemund)
> https://r4ds.had.co.nz/index.html

More free books!
> https://www.bigbookofr.com/

Shiny tutorial by R Studio
> https://shiny.rstudio.com/tutorial/written-tutorial/lesson1/

Want to get cooking? Join FoRTox!

Googling “how to (…) in R” will get you far, but still, it helps to have a community!



> Forensic Toxicology R Users Group

https://forms.gle/zejoxSemZGEHM7ak8

Acknowledgements

*Catherine Amirault *(LSJML)*

Prof. Félix Camirand Lemyre *(U. Sherbrooke)*

Cynthia Côté, M.Sc. *(LSJML)*

*Gabrielle Daigneault *(LSJML)*

*Alexandra Doyon *(LSJML)*

*Béatrice Garneau *(LSJML)*

*Marie-Jo Lajoie, M.Sc. *(LSJML)*

Julie Laquerre *(LSJML)*

Pascal Mireault, M.Sc. *(LSJML, U. Concordia)*

Marc-André Morel *(LSJML)*

Stéphanie Savard *(LSJML)*

Prof. Cameron D. Skinner *(U. Concordia)*

Questions? Comments?


brigitte.desharnais@msp.gouv.qc.ca

https://www.reasearchgate.net/profile/Brigitte-Desharnais

@ToxBrigitte

https://www.linkedin.com/in/brigitte-desharnais-ph-d-63624553

toxbrigitte

ToxBrigitte

GitHub: This is where you can find all my R scripts!

Find this presentation's material on GitHub

> https://github.com/ToxBrigitte/soft-2022-automation-workshop

- For those not familiar with GitHub, download everything on your computer like this: