

# ggplot2のアノテーション技法

Art of Annotation in ggplot2

24<sup>th</sup> February 2024, Tokyo.R #111

Yuta Kanzawa @yutakanzawa



Data Scientist at Zurich Insurance Company Limited, Japan Branch



# 神沢雄大 Yuta Kanzawa

- データサイエンティスト@チューリッヒ保険会社
  - 日本支店
- Twitter: [@yutakanzawa](https://twitter.com/yutakanzawa)
- 好きなもの：オペラとワイン
  - ワーグナー
  - ブルゴーニュ (WSET Lv 3→?)
- 使用可能言語：7
  - 人間：日本語、英語、ドイツ語
  - コンピューター：R, Python, SAS, SQL



# アジェンダ

- 今日話すこと
  - ggplot2によるアノテーション
  - ggforce
  - gghighlight
- 今日話さないこと
  - ggplot2そのもの

# おことわり

- 各パッケージの**公式ドキュメントの内容を抜粋した**ものです。
  - コード例は割愛。
  - 図表はことわりのない限り公式ドキュメントからの引用。
- **出力してから編集した方が手っ取り早い**こともあります。
  - プレゼンテーションソフト
  - 画像処理ソフト



# TL;DR

- 過剰なアノテーションに気をつける。
- ggplot2だけでアノテーション。 → **annotate()**
  - geom\_text()で文字が重なる。 → ggrepelパッケージ
- 散布図の一部の点を囲む。 → ggforceパッケージ
  - **geom\_mark\_xxx()**
- 一部の図形だけを強調する。 → **gghighlight**パッケージ
  - 順序に注意。

# アノテーションとは？

What's annotation?

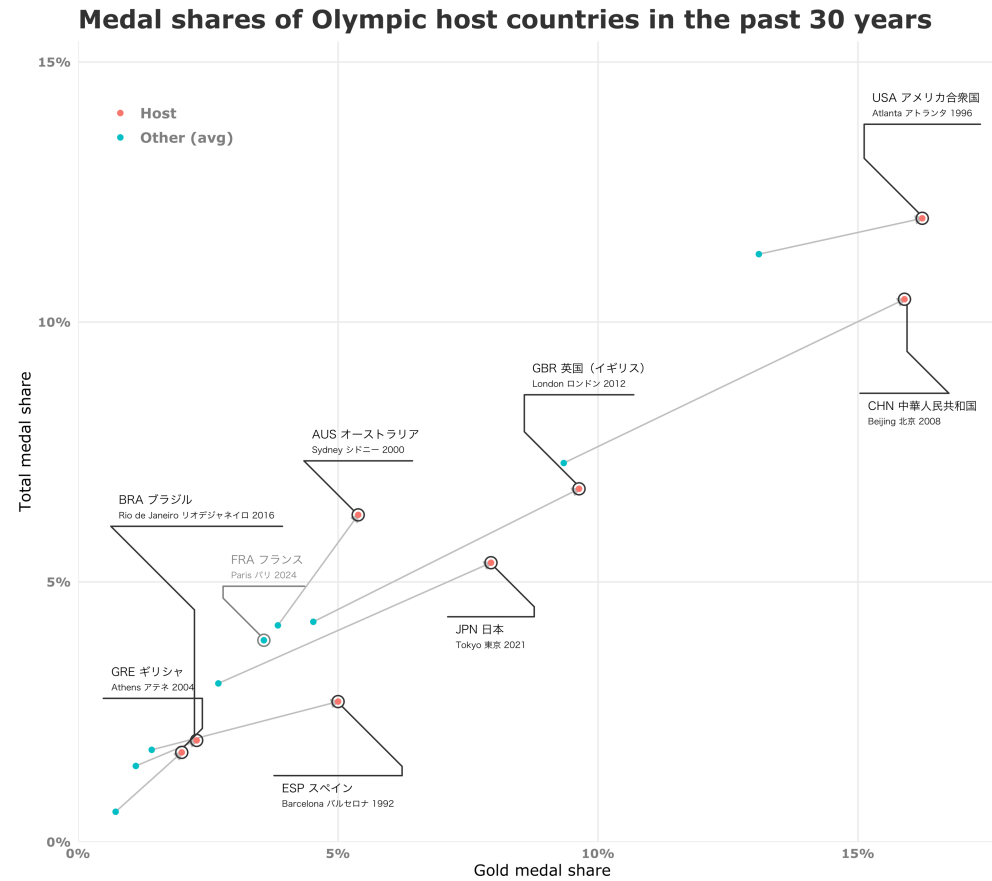
# [名]アノテーション【annotation】

- 本文につけた**注釈**。注解。
  - 出典：デジタル大辞泉（小学館）<sup>\*1</sup>
- あるデータに対して関連する情報（メタデータ）を**注釈として付与すること**。（中略）付与したメタデータやタグを指してアノテーションという場合もある。
  - 出典：Wikipedia<sup>\*2</sup>
- あくまで注釈なので**付け過ぎは禁物**！
  - 自省も込めて...

\*1 <https://dictionary.goo.ne.jp/word/%E3%82%A2%E3%83%8E%E3%83%86%E3%83%BC%E3%82%B7%E3%83%A7%E3%83%B3/>

\*2 <https://ja.wikipedia.org/wiki/%E3%82%A2%E3%83%8E%E3%83%86%E3%83%BC%E3%82%B7%E3%83%A7%E3%83%B3>

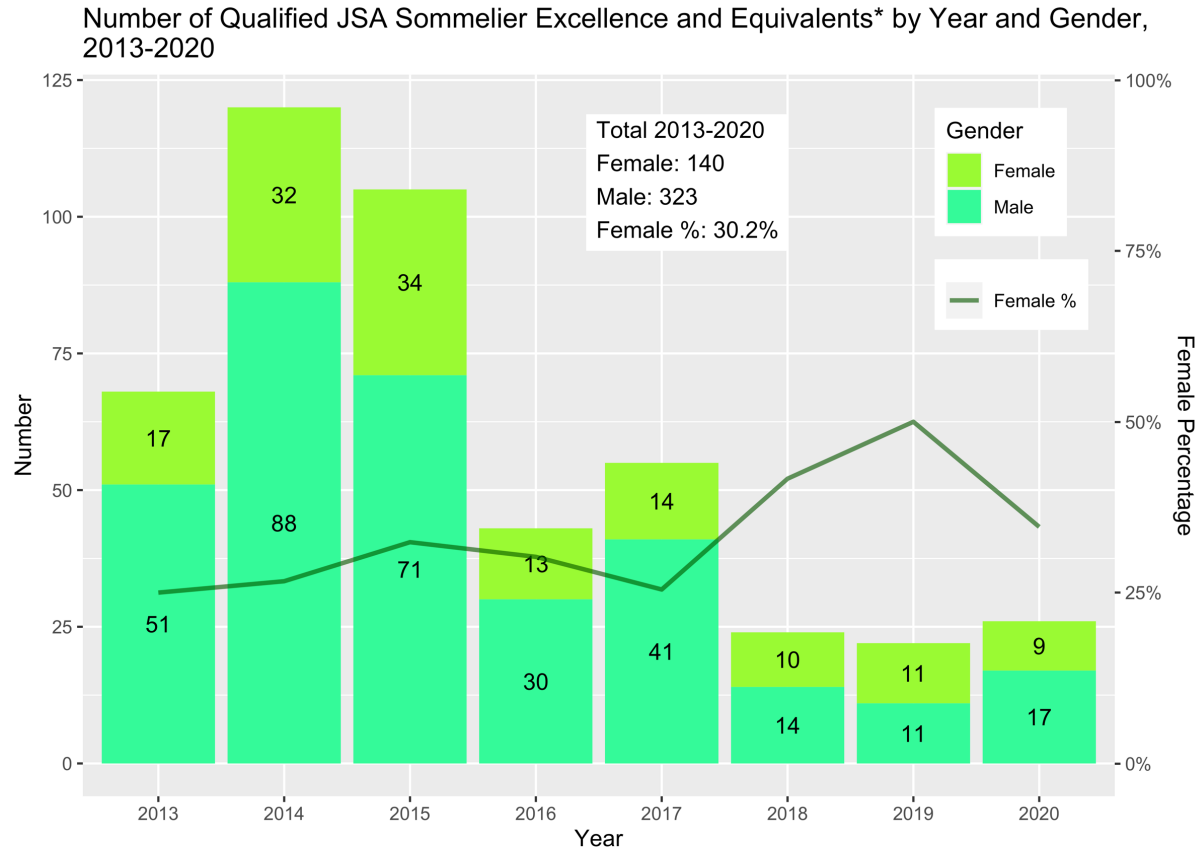
# 例1：散布図



Data: International Olympic Committee via <https://olympics.com> & <https://www.wikipedia.org> - Graphic: Yuta Kanzawa



# 例2：棒グラフ（1/2）

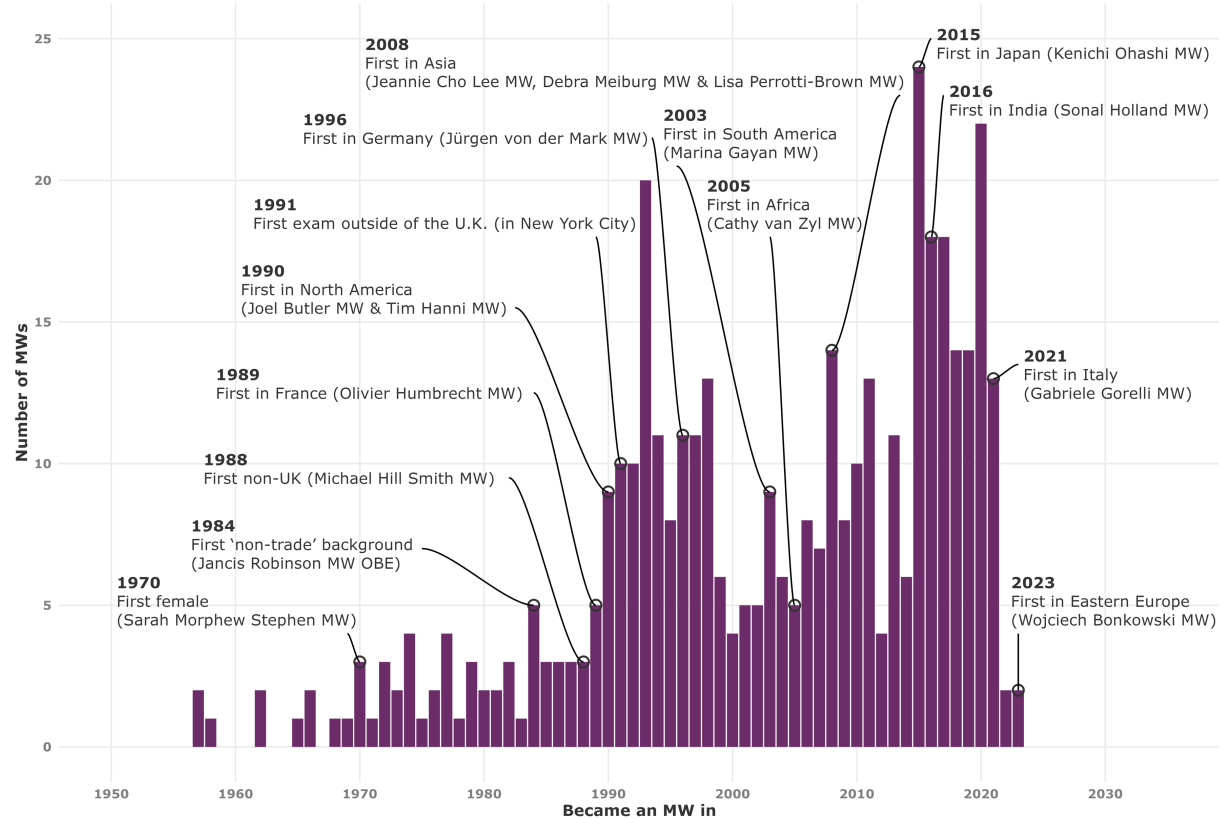


Source: Japan Sommelier Association <https://www.sommelier.jp/exam/pdf/qualifiedholders.pdf>  
\*Sommelier Excellence (2019-2020), Senior Sommelier (2013-2018), Senior Wine Adviser (2013-2015)

# 例3 : 棒グラフ (2/2)

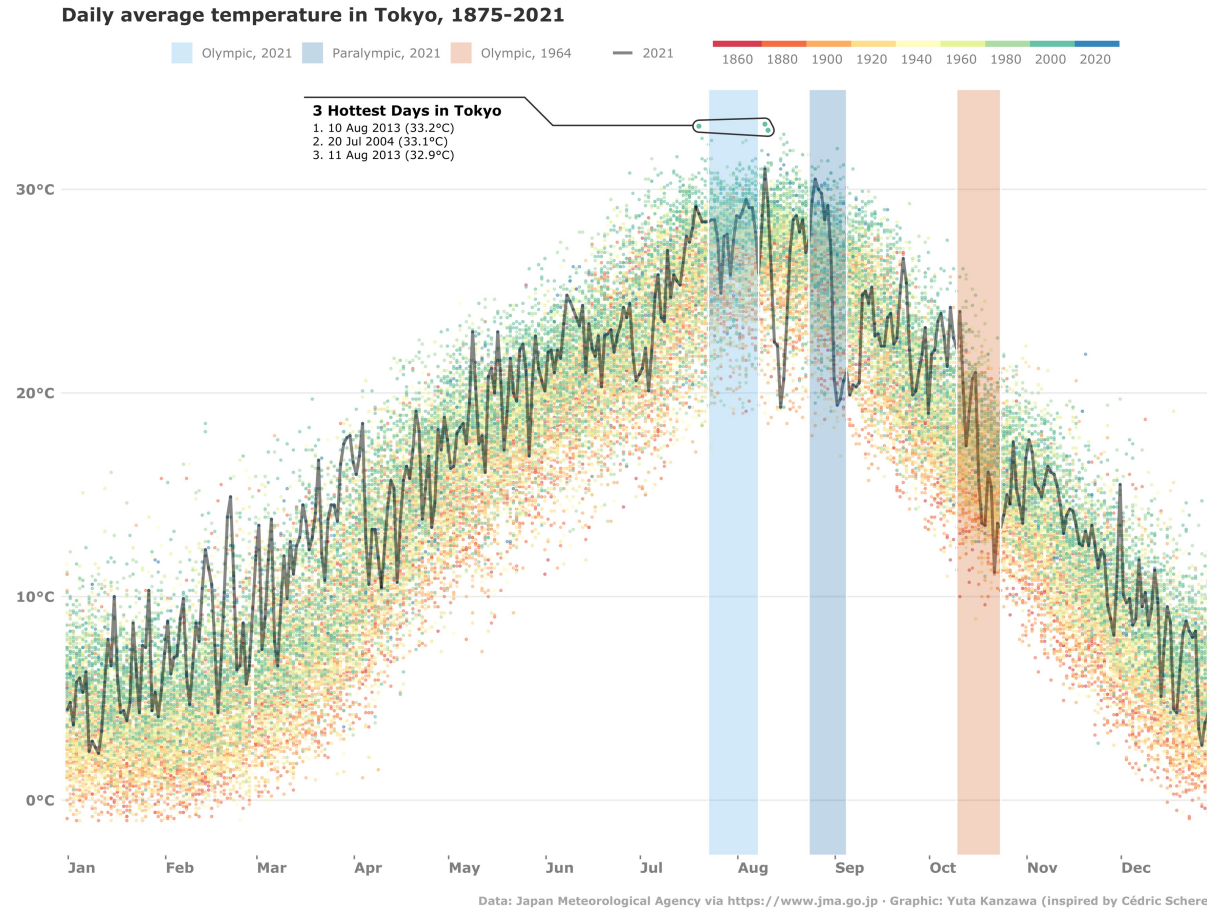
## 415 Active Masters of Wine by Year of Qualification

As of May 2023, 500 people have gained the title since the inaugural exam in May 1953. NB: 85 deceased or resigned MWs are not counted here.

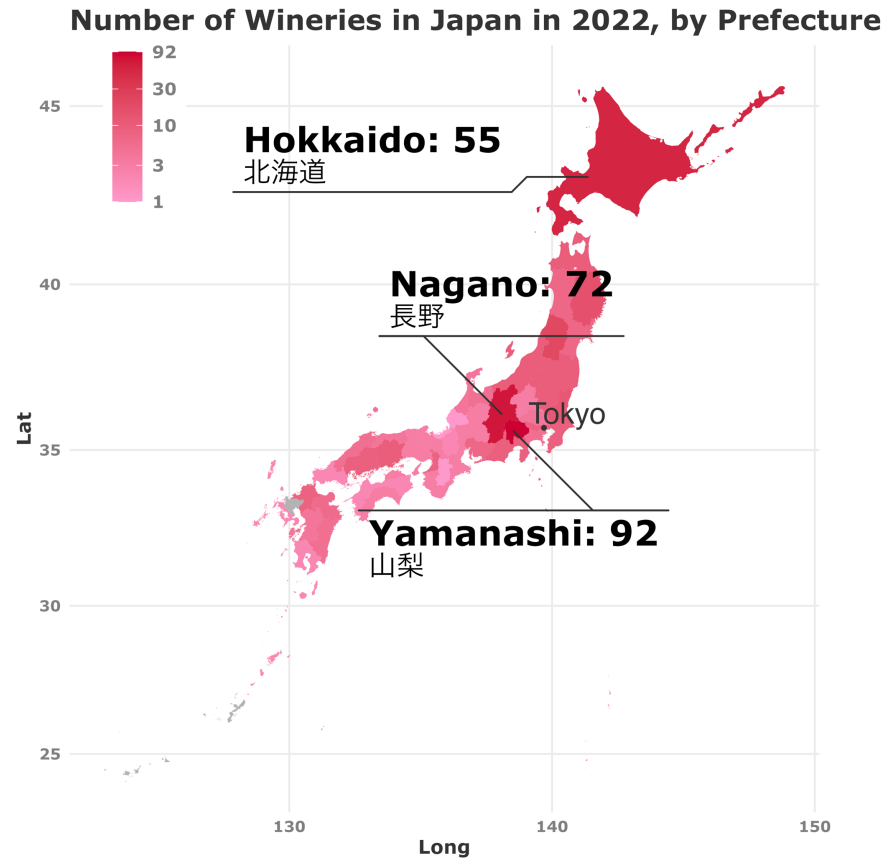


Data: The Institute of Masters of Wine via <https://www.mastersofwine.org> - Graphic: Yuta Kanzawa

# 例4：散布図 + 時系列



# 例5 : 地図



Data: National Tax Agency Japan via [https://www.nta.go.jp/taxes/sake/shiori-gaikyo/seizo\\_oroishiuri/r05/pdf/06.pdf](https://www.nta.go.jp/taxes/sake/shiori-gaikyo/seizo_oroishiuri/r05/pdf/06.pdf)  
Graphic: Yuta Kanzawa

# ggplot2

ggplot2



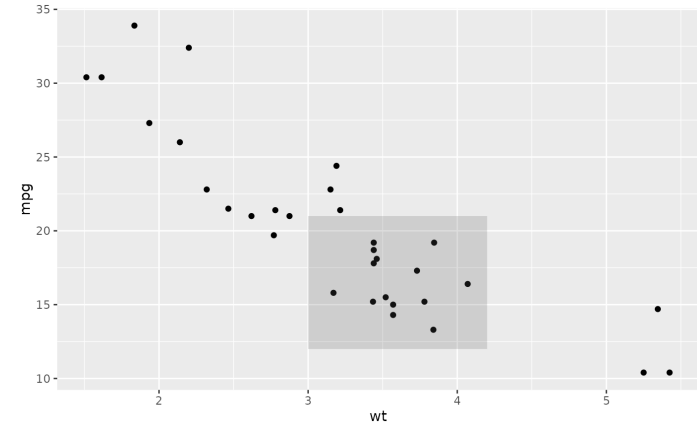
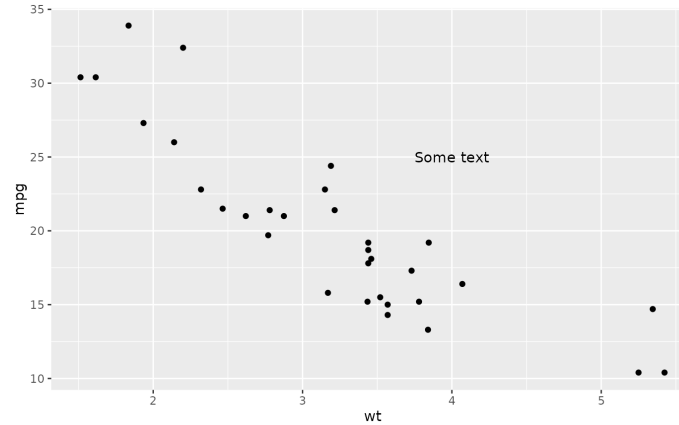
# geom\_text()? geom\_label()? annotate()?

- **アノテーション対象の違い**
  - **全てのデータ** → geom\_text() / geom\_label() → 2つの違いは後述。
  - **特定のデータ** → **annotate()** → 基本的にはこっち。
  - 出典：ggplot2の公式ドキュメント\*1
    - 'geom\_text() and geom\_label() add labels for each row in the data, [...]. To add labels at specified points use **annotate()** with annotate(geom = "text", ...) or annotate(geom = "label", ...).'
- その他アノテーションに関する公式FAQ
  - <https://ggplot2.tidyverse.org/articles/faq-annotation.html>

\*1 [https://ggplot2.tidyverse.org/reference/geom\\_text.html#details](https://ggplot2.tidyverse.org/reference/geom_text.html#details)

# annotate()

- geom関数との違い
  - aes()は引き継がれない。
  - データはベクトルで渡す。
  - **任意の位置**（座標や範囲）にアノテーションを配置可能。
- 文字以外のアノテーションも可能。
  - アノテーションに使うgeom関数名を第1引数geomに渡す。
    - 例：annotate(**geom = "text"**, ...)
  - ただし、abline、hline、vlineは使えない（仕様）\*1。
    - 当該geom関数を直接アノテーションに使用する。

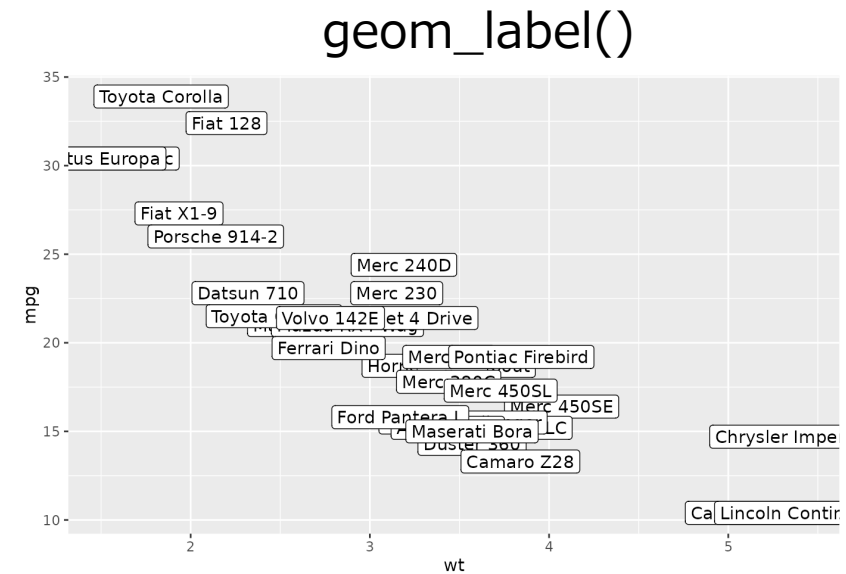
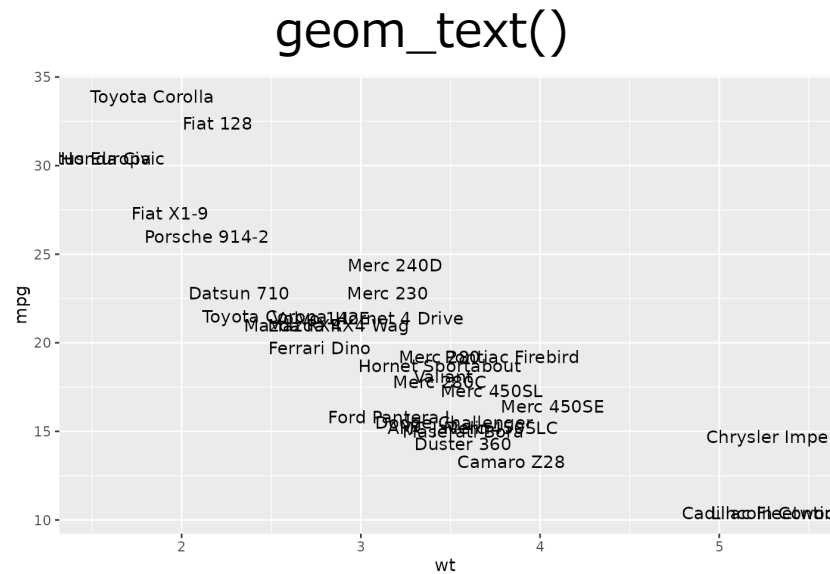


\*1 <https://ggplot2.tidyverse.org/reference/annotate.html#unsupported-geoms>

# geom\_text()とgeom\_label()

- 違い

- geom\_text() → **文字だけ**を表示。(左図)
- geom\_label() → 文字に**枠**を付ける。(右図)



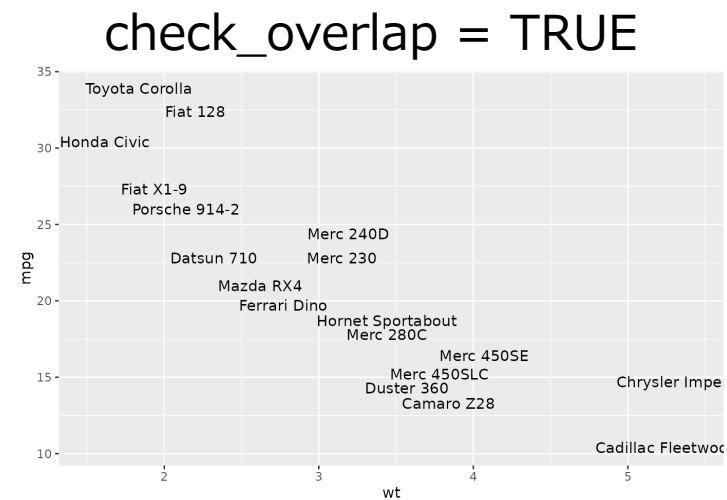
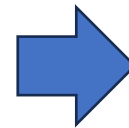
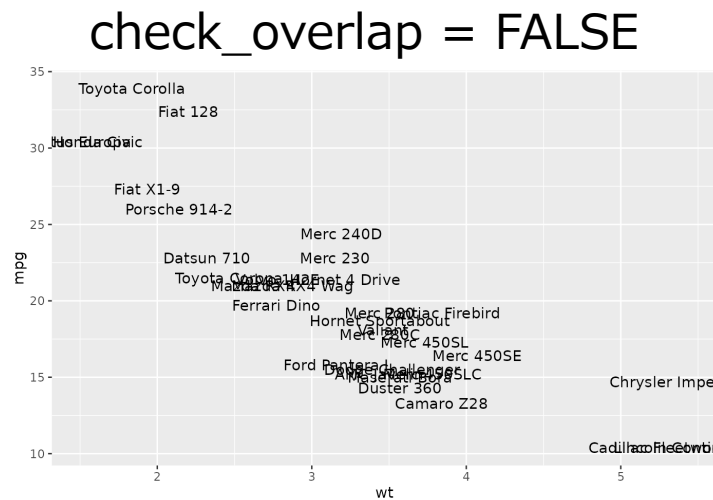
\*1 [https://ggplot2.tidyverse.org/reference/geom\\_text.html#details](https://ggplot2.tidyverse.org/reference/geom_text.html#details)



# geom\_text()とgeom\_label()で文字が重なるとき



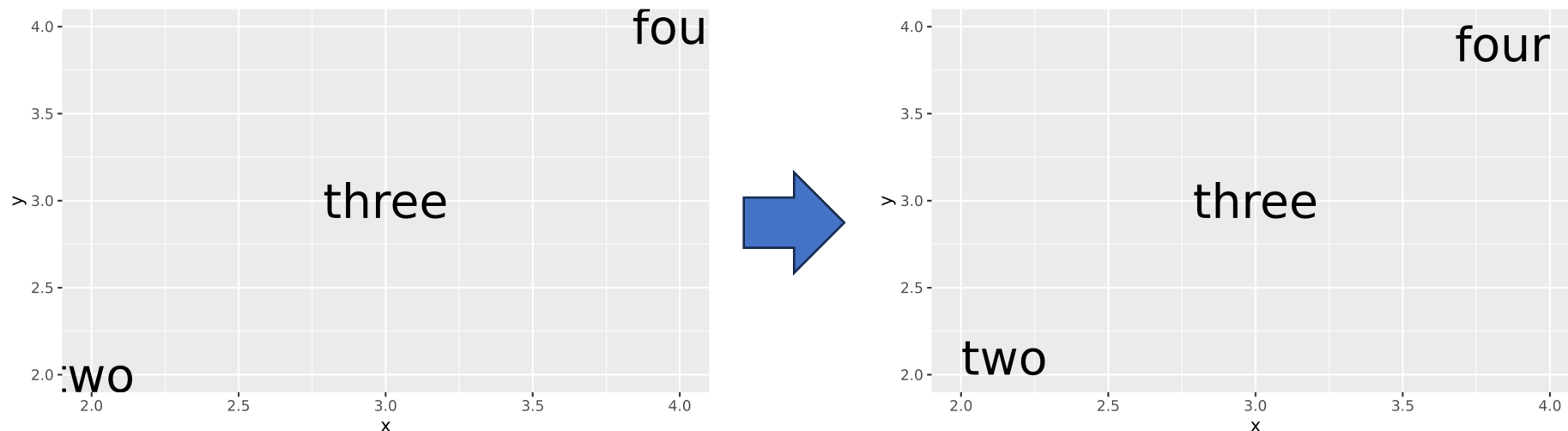
- 位置を調整して**全て描く**。 → **ggrepel**パッケージを使う\*1。
- 先に描いてあるものを優先。 → 引数**check\_overlap = TRUE**
  - **geom\_text()**のみ。
  - **データの順**に文字が描かれ、**既にあるものと重なる場合は描かれない**。



\*1 [https://ggplot2.tidyverse.org/reference/geom\\_text.html#details](https://ggplot2.tidyverse.org/reference/geom_text.html#details)

# geom\_text()とgeom\_label()で文字がはみ出すとき

- 引数**vjust**と**hjust**に**"inward"**を渡す\*1。
  - 文字の位置を描画範囲の中央に向かって寄せる。



\*1 [https://ggplot2.tidyverse.org/articles/faq-annotation.html?q=geom\\_text#how-can-i-make-sure-all-annotation-created-with-geom\\_text-fits-in-the-bounds-of-the-plot](https://ggplot2.tidyverse.org/articles/faq-annotation.html?q=geom_text#how-can-i-make-sure-all-annotation-created-with-geom_text-fits-in-the-bounds-of-the-plot)

# ggforce

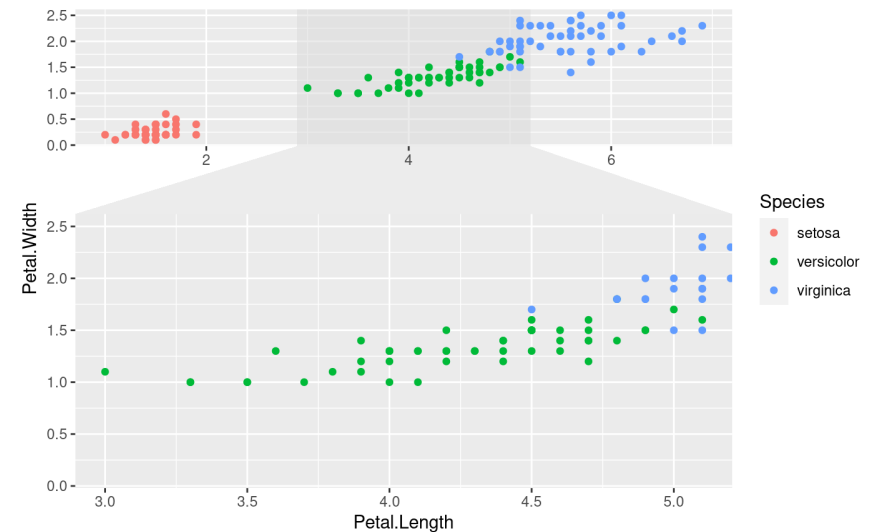
ggforce





# ggforceパッケージ

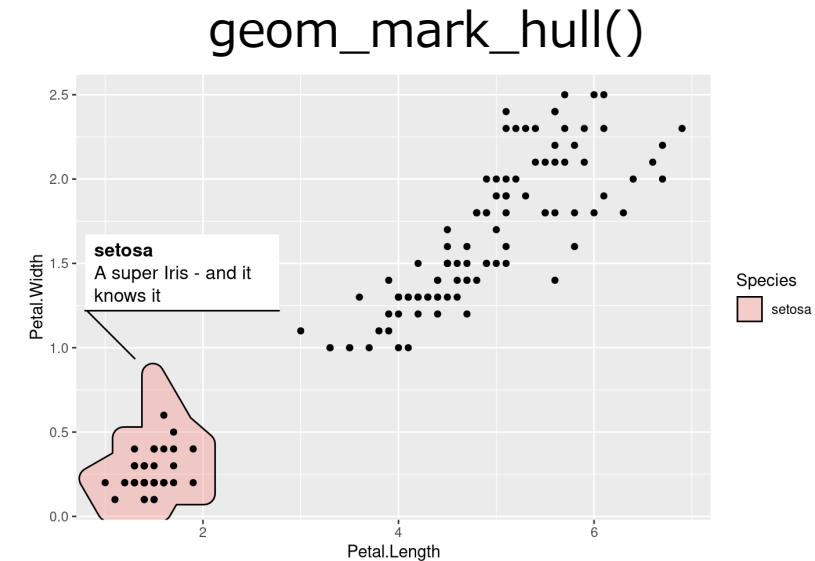
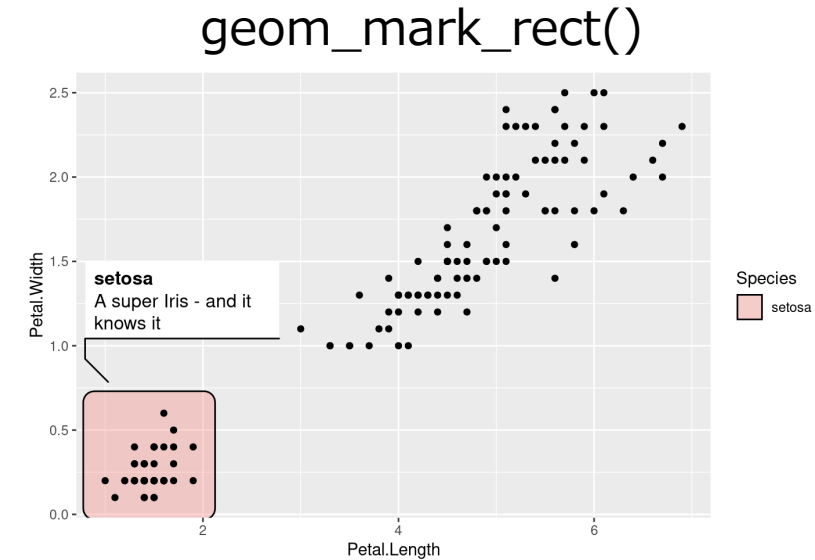
- ggplot2の機能を強化するためのパッケージ\*1
- アノテーション以外にも**便利な機能が豊富**。
  - 描画：円、楕円、ベジエ曲線など。
  - ファセット：複数カラム、拡大表示（右図）
  - 変換：指数
  - ラベル：TeX形式のファセットラベル
  - などなど。
- アノテーション：`geom_mark_xxx()`



\*1 <https://ggforce.data-imaginist.com/>

# geom\_mark\_xxx()

- 主に**散布図の一部の点を囲む**。
  - **見出しと説明**を指定可能。
  - 表示位置は内部で計算される（指定不可）。
- 囲み方
  - geom\_mark\_rect() → **四角**（上図）
  - geom\_mark\_circle() → 円
  - geom\_mark\_ellipse() → 楕円
  - geom\_mark\_hull() → **外郭**（下図）
    - 各点と外郭線の距離はパラメーターで調整可能\*1。



\*1 [https://ggforce.data-imaginist.com/reference/geom\\_mark\\_hull.html](https://ggforce.data-imaginist.com/reference/geom_mark_hull.html)

# gghighlight

gghighlight

# Tokyo.R第109回（2023年10月7日）：データ可視化特集

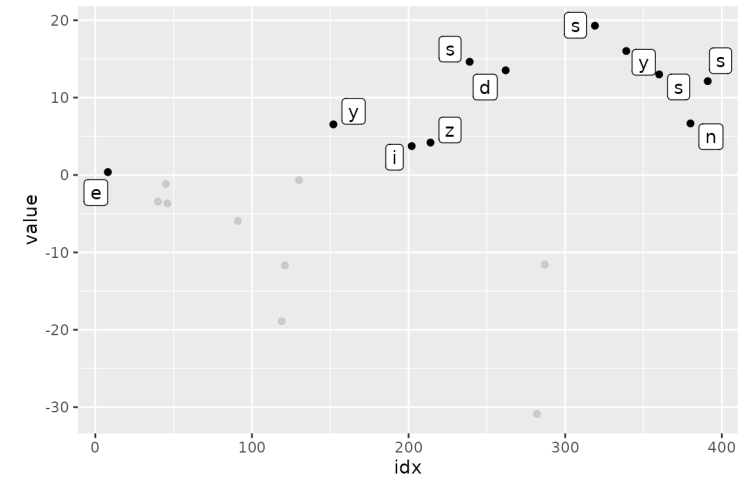
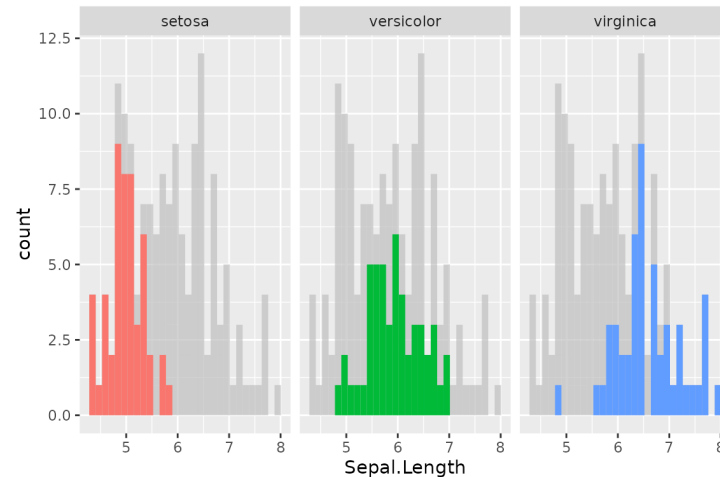
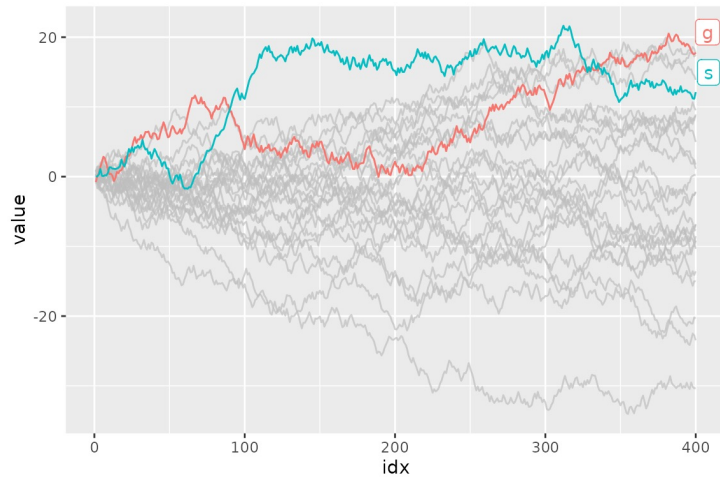
gghighlight の  
作者です。  
すべてを  
お話しします。

@yutannihilation

\* <https://tokyor.compass.com/event/296561/>

# gghighlightパッケージ

- 多数の図形から特定のもののだけをハイライトするパッケージ\*1
  - 折れ線グラフ、棒グラフ、散布図、ポリゴン（地理データ上の図形）など **ggplot2が扱えるものはほぼなんでも**。
  - 作者による丁寧&詳細な説明（前述） → ここでは要点のみ。



\*1 <https://yutannihilation.github.io/gghighlight/>



# gghighlight()

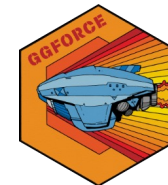
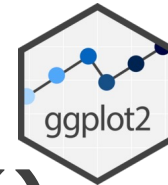
- **ハイライトする対象を特定する条件式**を第1引数に記述する。
  - `dplyr::filter()`がデータを抽出。
  - 例：
    - 折れ線グラフで最大値が20を超えるもの → `max(value) > 20`
    - 散布図で縦軸と横軸の値が共に正のもの → `value1 > 0 & value2 > 0`
      - 複数のAND条件を1つずつ引数に指定できるが、1つにまとめて論理関係を明示する方がよい。
        - `gghighlight(value1 > 0, value2 > 0, ...)` → `gghighlight(value1 > 0 & value2 > 0, ...)`
  - 実際には**グラフの重ね打ち**をしている。
    - `gghighlight()`は**最後に指定**するとよい。
    - ハイライト部分のスケールが元と異なる。 → **`keep_scale = TRUE`**

# まとめ

Long story short

# Long story short

- 過剰なアノテーションに気をつける。
- ggplot2だけでアノテーション。 → **annotate()**
  - geom\_text()で文字が重なる。 → ggrepelパッケージ
- 散布図の一部の点を囲む。 → ggforceパッケージ
  - geom\_mark\_xxx()
- 一部の図形だけを強調する。 → **gghighlight**パッケージ
  - 順序に注意。



# 参考

- *FAQ: Annotation*
  - ggplot2
    - <https://ggplot2.tidyverse.org/articles/faq-annotation.html>
- 『gghighlight の作者です。すべてをお話しします。』
  - yutannihilation
    - <https://yutani.quarto.pub/tokyor-109-gghighlight/>
- *ggplot Wizardry*
  - Dr. Cédric Scherer
    - [https://www.cedricscherer.com/slides/useR-2021\\_ggplot-wizardry-extended.pdf](https://www.cedricscherer.com/slides/useR-2021_ggplot-wizardry-extended.pdf)

# Enjoy!