

# A5. Compact Letter Display (CLD)

Buchstabenbasierte Zusammenfassung aller paarweisen Mittelwertvergleiche  
Dr. Paul Schmidt

Um alle in diesem Kapitel verwendeten Pakete zu installieren und zu laden, kann man folgenden Code ausführen:

```
for (pkg in c("emmeans", "multcomp", "multcompView", "ggtext", "scales",
"tidyverse")) {
  if (!require(pkg, character.only = TRUE)) install.packages(pkg)
}

library(emmeans)
library(multcomp)
library(multcompView)
library(ggtext)
library(scales)
library(tidyverse)
```

Ein Compact Letter Display (CLD) ist eine beliebte Methode, um die Ergebnisse aller paarweisen Vergleiche zwischen Behandlungsmittelwerten zusammenzufassen. Jeder Mittelwert wird mit einem oder mehreren Buchstaben versehen, und die Regel ist einfach: **Mittelwerte, die sich keinen Buchstaben teilen, unterscheiden sich signifikant.** Der zugrunde liegende Algorithmus wurde von H. P. Piepho, A. Büchse, and C. Richter [1] formalisiert, und seine Interpretation wurde von H.-P. Piepho [2] sorgfältig erneut beleuchtet.

Dieses Kapitel behandelt, wie man in R mit `emmeans` ein CLD erzeugt (der empfohlene moderne Arbeitsablauf), wie man die Buchstaben in eine `ggplot2`-Abbildung einbettet und wie man CLDs verantwortungsvoll interpretiert und berichtet. Wir setzen voraus, dass man mit Multiplizitätsadjustierungen bereits vertraut ist, wie sie im vorangegangenen Kapitel über multiple Vergleiche behandelt wurden.

## Die Kurzversion

Wir verwenden durchgängig den integrierten `PlantGrowth`-Datensatz. Der Arbeitsablauf ist immer derselbe: ein Modell fiten, mit `emmeans()` die geschätzten marginalen Mittelwerte berechnen und diese an `cld()` übergeben.

```
mod <- lm(weight ~ group, data = PlantGrowth)
mod_means <- emmeans(mod, specs = "group")
mod_means_cld <- cld(mod_means,
  adjust = "Tukey",
  Letters = letters,
  alpha = 0.05)
```

Note: `adjust = "tukey"` was changed to `"sidak"` because `"tukey"` is only appropriate for one set of pairwise comparisons

```
mod_means_cld
```

```

group emmean      SE df lower.CL upper.CL .group
trt1      4.66 0.197 27      4.16      5.16      a
ctrl      5.03 0.197 27      4.53      5.53      ab
trt2      5.53 0.197 27      5.02      6.03      b

```

Confidence level used: 0.95

Conf-level adjustment: sidak method for 3 estimates

P value adjustment: tukey method for comparing a family of 3 estimates

significance level used: alpha = 0.05

NOTE: If two or more means share the same grouping symbol,  
then we cannot show them to be different.  
But we also did not show them to be the same.

Die Spalte `.group` enthält die Buchstaben. Hier teilt sich `trt1` keinen Buchstaben mit `trt2`, weshalb sich diese beiden Mittelwerte auf dem 5%-Niveau signifikant unterscheiden. `ctrl` teilt sich mit beiden Buchstaben, sodass wir es von keinem der beiden unterscheiden können.

💡 Welches `cld()` rufen wir auf?

Es gibt im R-Ökosystem zwei `cld()`-Implementierungen: eine in `multcomp` (für `glht`-Objekte) und die Methode für `emmGrid`-Objekte, die `multcomp` über `emmeans` bereitstellt. Das Laden beider Pakete macht das Dispatching automatisch. Der alte direkte Aufruf `multcomp::cld(glht(...))` funktioniert weiterhin, ist aber nicht mehr der empfohlene Einstiegspunkt – `emmeans::emmeans() %>% cld()` ist flexibler, funktioniert für gemischte Modelle und berücksichtigt Multiplizitätsadjustierungen transparent.

## Die wichtigen Argumente

Die drei Argumente, die man am häufigsten anpasst, sind `adjust`, `Letters` und `alpha`.

- `adjust` steuert die Multiplizitätsadjustierung ( `"tukey"`, `"bonferroni"`, `"sidak"`, `"none"`, ...). Zur Theorie siehe das Kapitel über multiple Vergleiche (A4).
- `Letters = letters` erzwingt, dass die Ausgabe a-z verwendet. Die Voreinstellung ist numerisch, was man von einem sogenannten *Buchstaben*-Display selten möchte.
- `alpha` legt die Signifikanzschwelle dafür fest, ob zwei Mittelwerte getrennte Buchstaben erhalten.

i Eine verwirrende, aber harmlose Meldung

Man sieht möglicherweise einen Hinweis wie:

```

## Note: adjust = "tukey" was changed to "sidak"
## because "tukey" is only appropriate for one set of pairwise comparisons

```

Das ist kein Fehler. Die paarweisen *p*-Werte sind weiterhin Tukey-adjustiert; lediglich die Konfidenzintervalle der Mittelwerte selbst fallen auf die Sidak-Adjustierung zurück, da Tukey für Kontraste definiert ist, nicht für einzelne Mittelwerte. Hintergrund: [stats.stackexchange.com](https://stats.stackexchange.com).

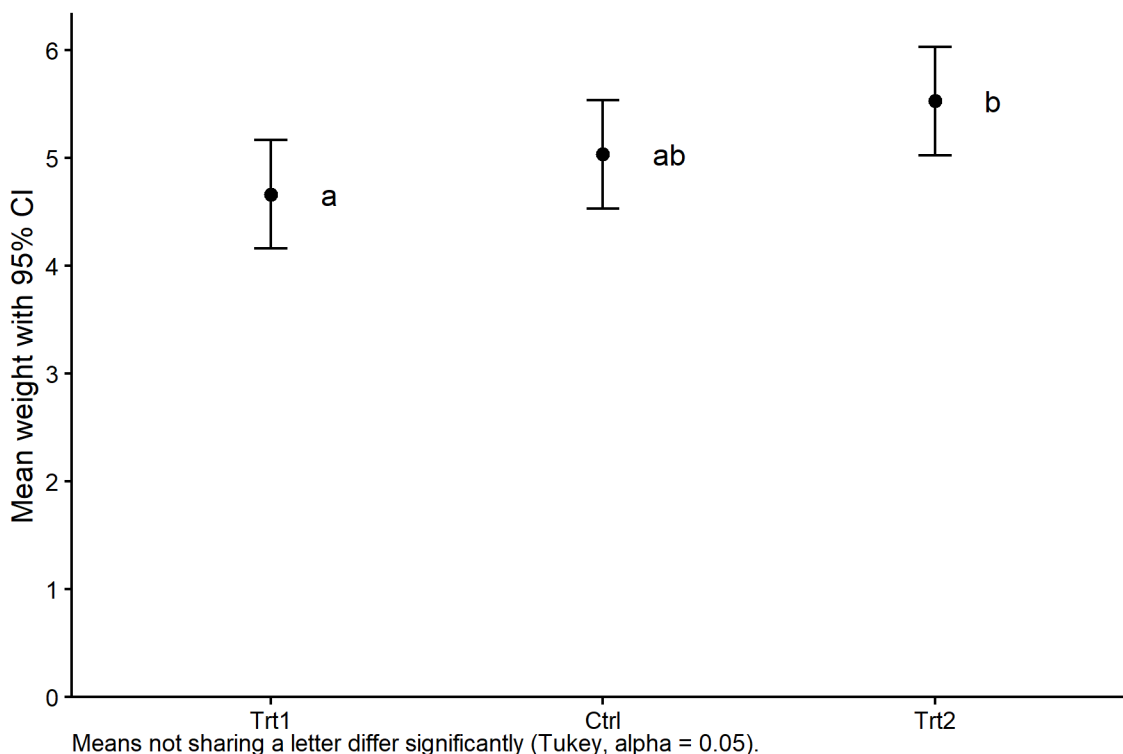
# Das CLD plotten

Die Buchstaben in einen Plot einzubauen, ist die häufigste Folgefrage. Das Vorgehen ist:

1. Die `cld()`-Ausgabe in ein Tibble umwandeln.
2. Leerzeichen in `.group` entfernen (`cld()` füllt die Buchstaben zur Ausrichtung mit Leerzeichen auf).
3. Die bereinigten Buchstaben auf eine Text-Ebene mappen.

```
mod_means_cld_tbl <- mod_means_cld %>%
  as_tibble() %>%
  mutate(.group = str_trim(.group)) %>%
  mutate(group = fct_reorder(group, emmean))

ggplot(mod_means_cld_tbl, aes(x = group, y = emmean)) +
  geom_point(size = 2) +
  geom_errorbar(aes(ymin = lower.CL, ymax = upper.CL), width = 0.1) +
  geom_text(aes(label = .group), nudge_x = 0.15, hjust = 0) +
  scale_y_continuous(
    name = "Mean weight with 95% CI",
    limits = c(0, NA),
    breaks = pretty_breaks(),
    expand = expansion(mult = c(0, 0.05))
  ) +
  scale_x_discrete(name = NULL, labels = str_to_title) +
  labs(caption = "Means not sharing a letter differ significantly (Tukey, alpha =
0.05).") +
  theme_classic() +
  theme(plot.caption = element_textbox_simple(hjust = 1))
```



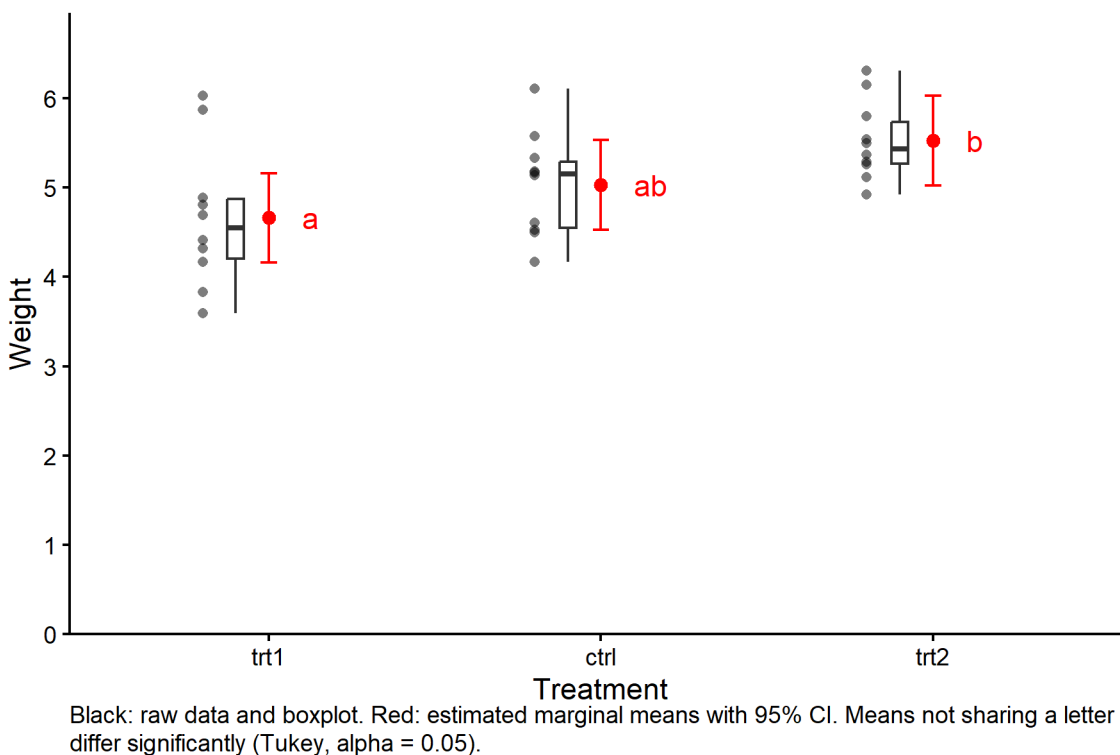
Für anspruchsvolleres Styling (Schriftarten, Farbpaletten, Theme-Komponenten) siehe die Kapitel in der `ggplot2`-Sektion dieser Website.

## Ein reichhaltigerer Plot

Der minimale Plot oben fasst die Modellausgabe zusammen. Manchmal lohnt es sich, die Rohdaten zusammen mit den geschätzten Mittelwerten zu zeigen, damit die Lesenden sowohl die Effektstärke als auch die Datenqualität auf einen Blick beurteilen können:

```
PlantGrowth2 <- PlantGrowth %>%
  mutate(group = fct_relevel(group, levels(mod_means_cld_tbl$group)))

ggplot() +
  geom_point(data = PlantGrowth2,
            aes(x = group, y = weight),
            shape = 16, alpha = 0.5,
            position = position_nudge(x = -0.2)) +
  geom_boxplot(data = PlantGrowth2,
             aes(x = group, y = weight),
             width = 0.05, outlier.shape = NA,
             position = position_nudge(x = -0.1)) +
  geom_point(data = mod_means_cld_tbl,
            aes(x = group, y = emmean),
            size = 2, color = "red") +
  geom_errorbar(data = mod_means_cld_tbl,
              aes(x = group, ymin = lower.CL, ymax = upper.CL),
              width = 0.05, color = "red") +
  geom_text(data = mod_means_cld_tbl,
           aes(x = group, y = emmean, label = .group),
           nudge_x = 0.1, hjust = 0, color = "red") +
  scale_y_continuous(name = "Weight",
                    limits = c(0, NA),
                    breaks = pretty_breaks(),
                    expand = expansion(mult = c(0, 0.1))) +
  scale_x_discrete(name = "Treatment") +
  labs(caption = glue::glue(
    "Black: raw data and boxplot. Red: estimated marginal means with 95% CI. ",
    "Means not sharing a letter differ significantly (Tukey, alpha = 0.05).")
  )) +
  theme_classic() +
  theme(plot.caption = element_textbox_simple())
```



Balkendiagramme mit Fehlerbalken (“Dynamite Plots”) sind eine weitere traditionelle Wahl, sie verbergen jedoch die zugrunde liegende Verteilung und werden in der modernen Datenvisualisierung generell nicht empfohlen.

## CLDs sorgfältig interpretieren

CLDs sind praktisch, doch sie bringen eine gut dokumentierte Fallstrick mit sich: **Sie berichten Nicht-Befunde, keine Befunde.** Dass sich zwei Mittelwerte einen Buchstaben teilen, *beweist nicht*, dass sie gleich sind – es bedeutet lediglich, dass der Test die Gleichheit nicht verwerfen konnte. H.-P. Piepho [2] macht diesen Punkt explizit und empfiehlt eine einzige, eindeutige Bildunterschrift:

Mittelwerte, die keinen Buchstaben teilen, unterscheiden sich auf dem gewählten Signifikanzniveau signifikant.

Man sollte Formulierungen wie “Mittelwerte mit demselben Buchstaben sind gleich” oder “Mittelwerte mit demselben Buchstaben unterscheiden sich nicht” vermeiden. Beide überzeichnen, was die Buchstaben tatsächlich aussagen.

### ! Russell Lenth's Warnung

Der Maintainer von `emmeans`, Russell V. Lenth, hat sich deutlich zu den Grenzen von CLDs geäußert. Zwei häufige Kritikpunkte:

- **Schwarz-Weiß-Schwellenwertbildung.** Ein p-Wert von 0,049 und einer von 0,051 sind praktisch identisch, doch ein CLD behandelt sie als Gegensätze.
- **Verschleierung von Effektstärken.** Die Buchstaben verdichten eine reichhaltige Menge paarweiser Vergleiche zu einer einzigen kategorialen Zusammenfassung.

Das Paket gab früher sogar den Hinweis *“Compact letter displays can be misleading because they show NON-findings rather than findings.”* aus. Die aktuelle Formulierung ist milder, transportiert aber dieselbe Botschaft: *“If two or more means share the same grouping letter, then we cannot show them to be different. But we also did not show them to be the same.”* Siehe Lenth's Kommentar auf Cross Validated.

Wer Effektstärken transparenter kommunizieren möchte, sollte erwägen, die paarweisen Vergleiche direkt zu zeigen (nächster Abschnitt) - zusätzlich zum CLD oder anstelle davon.

## Die zugrunde liegenden Vergleiche betrachten

Die Buchstaben werden aus paarweisen Vergleichen aufgebaut. Um diese Vergleiche zu inspizieren, verwendet man `pairs()` mit `infer = c(TRUE, TRUE)`, um sowohl Konfidenzintervalle als auch p-Werte zu erhalten:

```
mod_means %>%
  pairs(adjust = "tukey", infer = c(TRUE, TRUE))
```

contrast	estimate	SE	df	lower.CL	upper.CL	t.ratio	p.value
ctrl - trt1	0.371	0.279	27	-0.32	1.062	1.331	0.3909
ctrl - trt2	-0.494	0.279	27	-1.19	0.197	-1.772	0.1980
trt1 - trt2	-0.865	0.279	27	-1.56	-0.174	-3.103	0.0120

```
Confidence level used: 0.95
Conf-level adjustment: tukey method for comparing a family of 3 estimates
P value adjustment: tukey method for comparing a family of 3 estimates
```

Alternativ liefert die Übergabe von `details = TRUE` an `cld()` sowohl das Buchstaben-Display als auch die Vergleiche in einem einzigen Objekt zurück:

```
cld(mod_means, adjust = "tukey", Letters = letters, details = TRUE)
```

```
Note: adjust = "tukey" was changed to "sidak"
because "tukey" is only appropriate for one set of pairwise comparisons
```

```
$emmeans
  group emmean      SE df lower.CL upper.CL .group
  trt1  4.66 0.197 27    4.16    5.16    a
  ctrl  5.03 0.197 27    4.53    5.53   ab
  trt2  5.53 0.197 27    5.02    6.03    b

Confidence level used: 0.95
Conf-level adjustment: sidak method for 3 estimates
P value adjustment: tukey method for comparing a family of 3 estimates
significance level used: alpha = 0.05
NOTE: If two or more means share the same grouping symbol,
      then we cannot show them to be different.
      But we also did not show them to be the same.

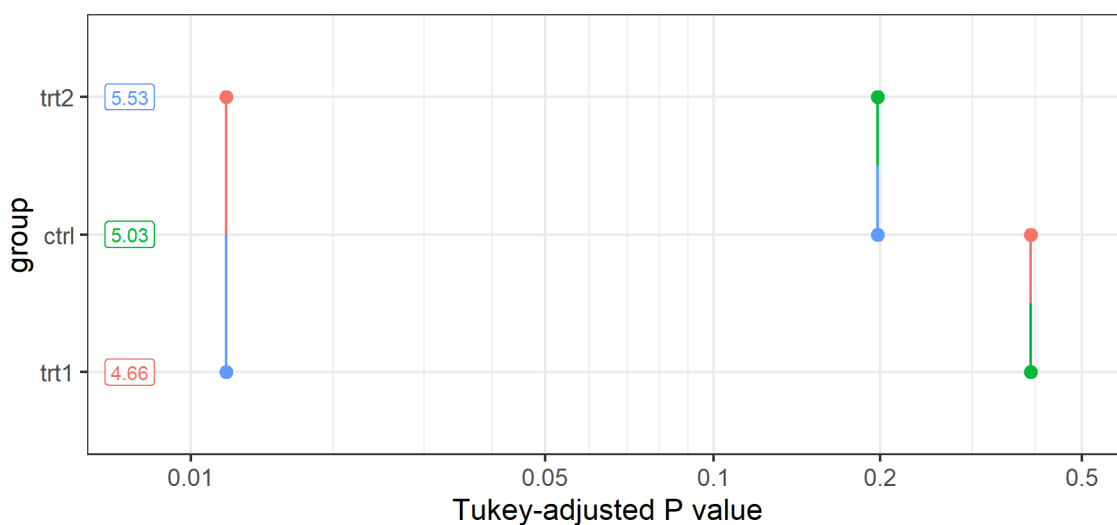
$comparisons
  contrast      estimate      SE df t.ratio p.value
  ctrl - trt1    0.371 0.279 27   1.331 0.3909
  trt2 - trt1    0.865 0.279 27   3.103 0.0120
  trt2 - ctrl    0.494 0.279 27   1.772 0.1980

P value adjustment: tukey method for comparing a family of 3 estimates
```

## Alternativen zum CLD

`emmeans` stellt zwei visuelle Alternativen bereit, die die Fallstricke des Buchstaben-Displays vermeiden:

```
pwpp(mod_means) + theme_bw()
```



Der **paarweise p-Wert-Plot** (`pwpp()`) zeigt jeden p-Wert auf einer horizontalen Achse und verbindet die beiden verglichenen Stufen mit einem Liniensegment. Das macht p-Werte als Kontinuum sichtbar und nicht als binäre Signifikanzentscheidung.

Die **paarweise p-Wert-Matrix** (`pwpm()`) gibt eine kompakte Matrix mit p-Werten oberhalb der Diagonalen und geschätzten Differenzen unterhalb aus und ist nützlich in Berichten, in denen eine Abbildung übertrieben wäre.

## Wann ein CLD verwenden (und wann nicht)

Ein CLD ist eine pragmatische Zusammenfassung. Es ist angebracht, wenn:

- Man eine moderate Anzahl an Behandlungsstufen hat (sagen wir bis etwa zehn).
- Das Publikum buchstabenbasierte Zusammenfassungen erwartet oder mit ihnen vertraut ist.
- Eine einzige Tabelle oder ein Plot viele Vergleiche auf einmal vermitteln muss.

Weniger geeignet ist es, wenn:

- Effektstärken wichtiger sind als ein Ja/Nein-Signifikanzurteil (Konfidenzintervalle zeigen).
- Die Anzahl der Mittelwerte groß ist – der Buchstabencode wird dann schwer lesbar.
- Der Kontext einer ist, in dem das Teilen eines Buchstabens als “äquivalent” missverstanden werden könnte statt als “nicht als unterschiedlich nachgewiesen”.

### i Die wichtigsten Punkte

1. Man verwendet `emmeans() %>% cld()` als Standard-Arbeitsablauf. Dabei setzt man `Letters = letters` und wählt eine `adjust`-Methode bewusst aus.
2. Die Buchstaben kodieren Nicht-Befunde, keine Befunde. Bildunterschriften von Plots sollten entsprechend formuliert werden.
3. CLDs ergänzen die zugrunde liegenden paarweisen Vergleiche, ersetzen sie aber nicht. `pairs()` und `pwpp()` bleiben wertvolle diagnostische Werkzeuge.
4. Man entfernt Leerzeichen in `.group` vor dem Plotten und bevorzugt Punkt- und Fehlerbalken-Plots gegenüber Balkendiagrammen für die Buchstabenbeschriftung.

## Bibliography

- [1] H. P. Piepho, A. Bückle, and C. Richter, “A Mixed Modelling Approach for Randomized Experiments with Repeated Measures,” *Journal of Agronomy and Crop Science*, vol. 190, no. 4, pp. 230–247, 2004, doi: 10.1111/j.1439-037X.2004.00097.x.
- [2] H.-P. Piepho, “Letters in Mean Comparisons: What They Do and Don’t Mean,” *Agronomy Journal*, vol. 110, no. 2, pp. 431–434, Mar. 2018, doi: 10.2134/agronj2017.10.0580.