Applied R Munich: "A Grammar of Data Manipulation" – Eine Einführung in das Paket dplyr
Tutorial

Philipp J. Rösch

26.10.2015

Die Fragen stammen größtenteils aus Hadley Wickhams dplyr-Tutorial von der useR! 2014. Danke hierfür!

```
suppressMessages(library(dplyr))
library(hflights)
flights <- tbl_df(hflights)
```

1. Welcher Flug hat am meisten Verspätung aufgeholt? Verwende `FlightNum`, `DepDelay` und `ArrDelay`.

2. Berechne die Geschwindigkeit in mph mit `AirTime` (in Minuten) und `Distance` (in miles). Erstelle außerdem die Variable Geschwindigkeit in km/h. Welche Flugzeuge (`FlightNum`) sind am schnellsten? Zusatzfrage: Erstelle eine Häufigkeitstabelle der Airlines für die 20 schnellsten Flüge.

3. Welche Airline ist im Durchschnitt am schnellsten?

```
flights4 <- flights %>%
  mutate(hour = DepTime %/% 100, date = sprintf("%04s-%02s-%02s", Year, Month, DayofMonth))

hourly_delay <- filter(
  summarise(
    group_by(
      filter(
        flights4, !is.na(DepDelay)
      ),
      date, hour
    ),
    avg_delay = mean(DepDelay),
    n = n()
  ),
  n > 10
)
```

4. Schreibe den oben stehenden Code in die Chaining-Syntax um.

5. Um wie viel Uhr starten jeweils täglich die ersten Flieger vom George Bush Intercontinental Airport (IAH)? `DepTime` ist hier ein Integer.

6. Gebe für jedes Flugzeug die zwei Flüge mit der meisten Verspätung aus. Was ist hier der Unterschied zwischen `min_rank`, `row_number` und `dense_rank`? Zusatzaufgabe: Welche Flüge gab es bloß einmal in 2011?