

# Reproduceerbaar analyseren voor praktijkgerichte onderzoekers:

Dag 4



# Samenvatting: tibbles / data.frames



	key_ID	village	interview_date	no_membrs
1	1	God	2016-11-17	3
2	1	God	2016-11-17	7
3	3	God	2016-11-17	10
4	4	God	2016-11-17	7
5	5	God	2016-11-17	7
6	6	God	2016-11-17	3

# Samenvatting: dplyr en tidyr

- **select()**: subset columns
- **filter()**: subset rows on conditions
- **mutate()**: calculate stuff

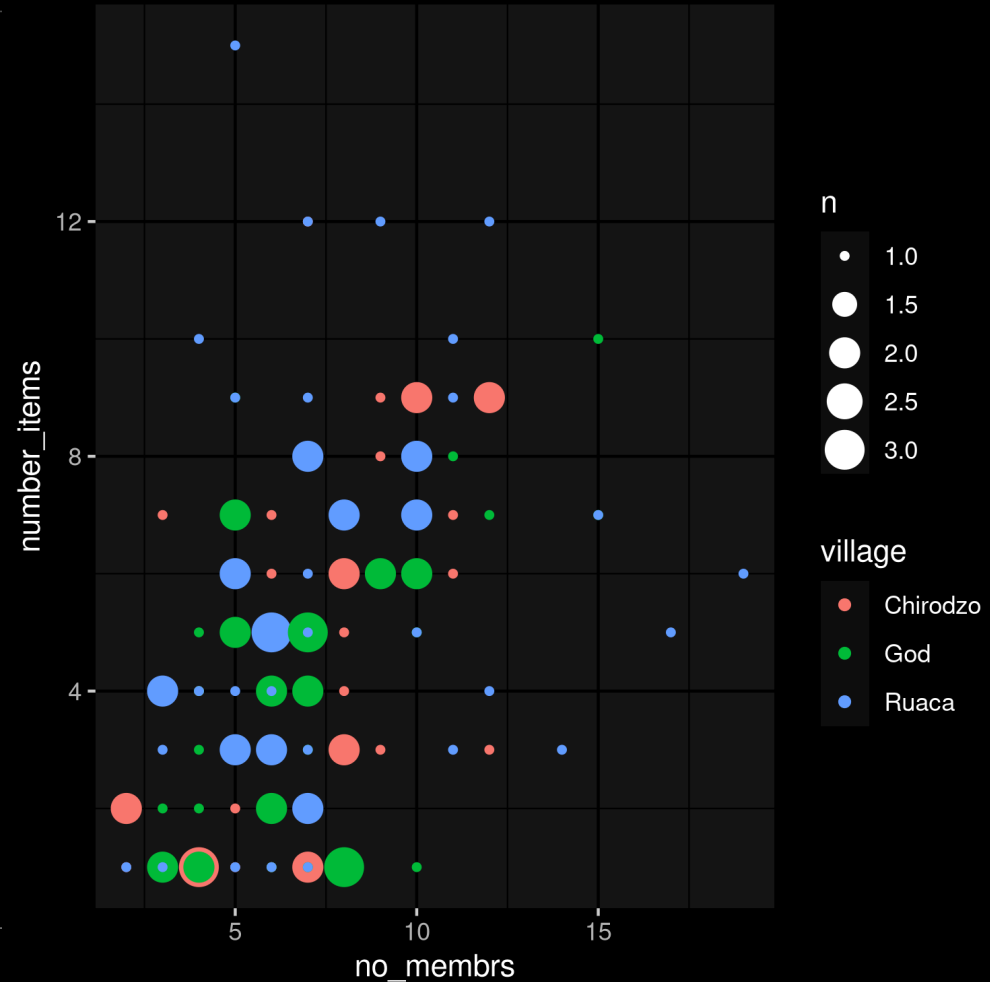
```
> interviews %>%  
+   mutate(total_meals = no_membrs * no_meals) %>%  
+   filter(total_meals > 20) %>%  
+   select(village, total_meals) %>%  
+   head()  
# A tibble: 6 × 2  
  village    total_meals  
  <chr>         <dbl>  
1 Chirodzo         24  
2 Chirodzo         24  
3 Chirodzo         36  
4 God              21  
5 God              30  
6 God              27
```

# Vandaag: dplyr en tidyr vervolg

- **select()**: subset columns
- **filter()**: subset rows on conditions
- **mutate()**: calculate stuff
- **group\_by()** & **summarize()**: summary statistics
- **arrange()**: sort results
- **count()**: count discrete values

# Vandaag: ggplot

- Componenten van een ggplot
- Scatter-, box-, en barplots
- Aesthetics (bijv. kleur)
- Meerdere plots



# Opdracht: dplyr en tidyr

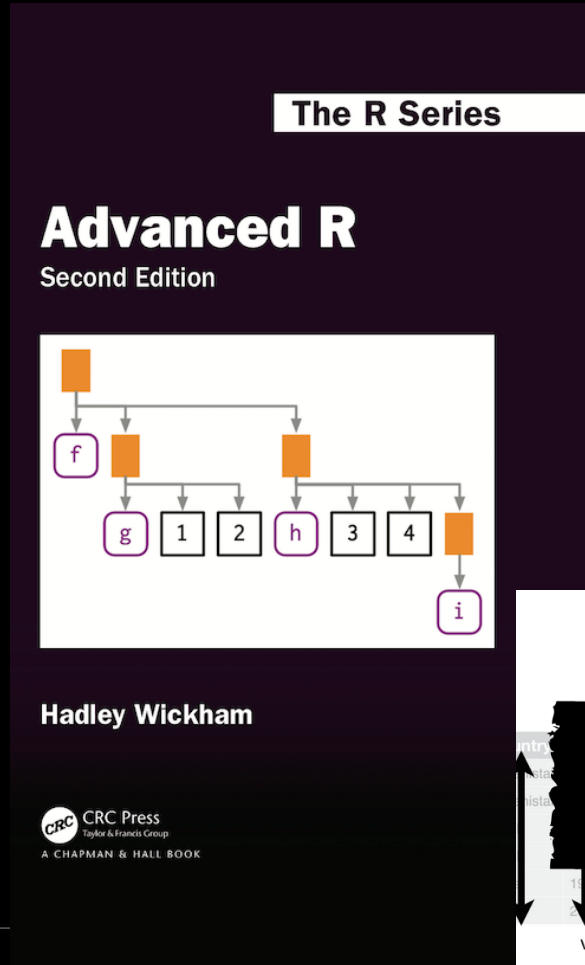
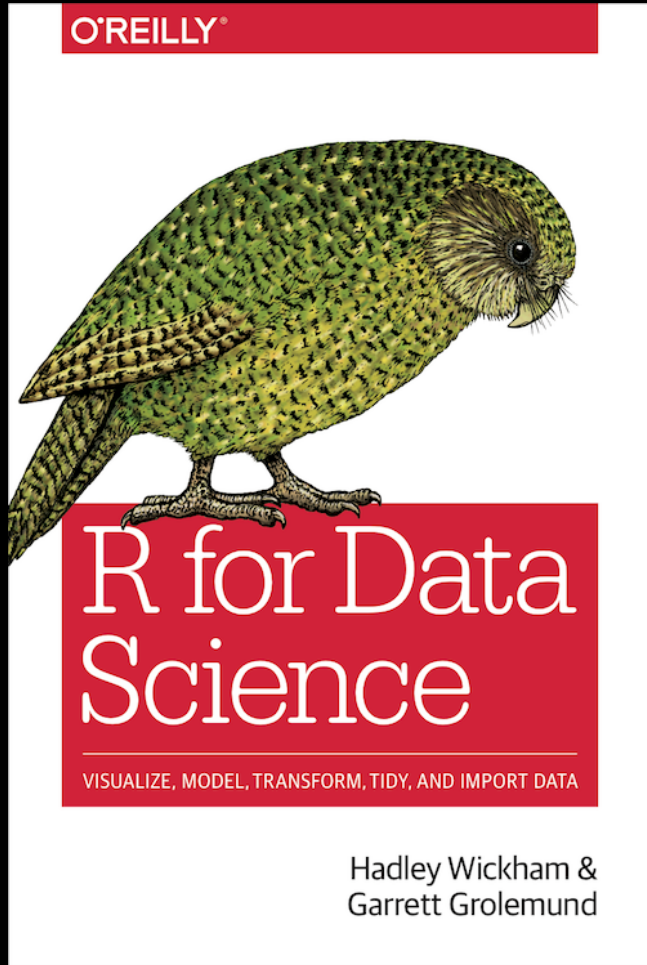
- How many households in the survey have an average of two meals per day? Three meals per day? Are there any other numbers of meals represented?
- Use `group_by()` and `summarize()` to find the mean, min, and max number of household members for each village. Also add the number of observations (hint: see ?n).
- What was the largest household interviewed in each month?

# Opdracht: ggplot

---

Use what you just learned to create a scatter plot of rooms by village with the `respondent_wall_type` showing in different colours. Does this seem like a good way to display the relationship between these variables? What other kinds of plots might you use to show this type of data?

# En nu?



# R-Ladies

