

Tidying Data

(dimensions change)

`slides/11-Tidying.pdf`

Tidy data

Hadley Wickham, "Tidy Data", *Journal of Statistical Software* 59 (10), August 2014.

- <http://vita.had.co.nz/papers/tidy-data.pdf>
- <https://www.jstatsoft.org/article/view/v059i10>

Messy 1

	treatmenta	treatmentb
John Smith	—	2
Jane Doe	16	11
Mary Johnson	3	1

Messy 2

	John Smith	Jane Doe	Mary Johnson
treatmenta	—	16	3
treatmentb	2	11	1

Tidy

name	trt	result
John Smith	a	—
Jane Doe	a	16
Mary Johnson	a	3
John Smith	b	2
Jane Doe	b	11
Mary Johnson	b	1

messy

	id	city	hwy
1	car1	19	24
2	car2	20	30
3	car3	29	35

tidy

	id	roadtype	mpg
1	car1	city	19
2	car2	city	20
3	car3	city	29
4	car1	hwy	24
5	car2	hwy	30
6	car3	hwy	35

For more detail, see: "Tidyr::gather()" tutorial
www.github.com/jtr13/codehelp

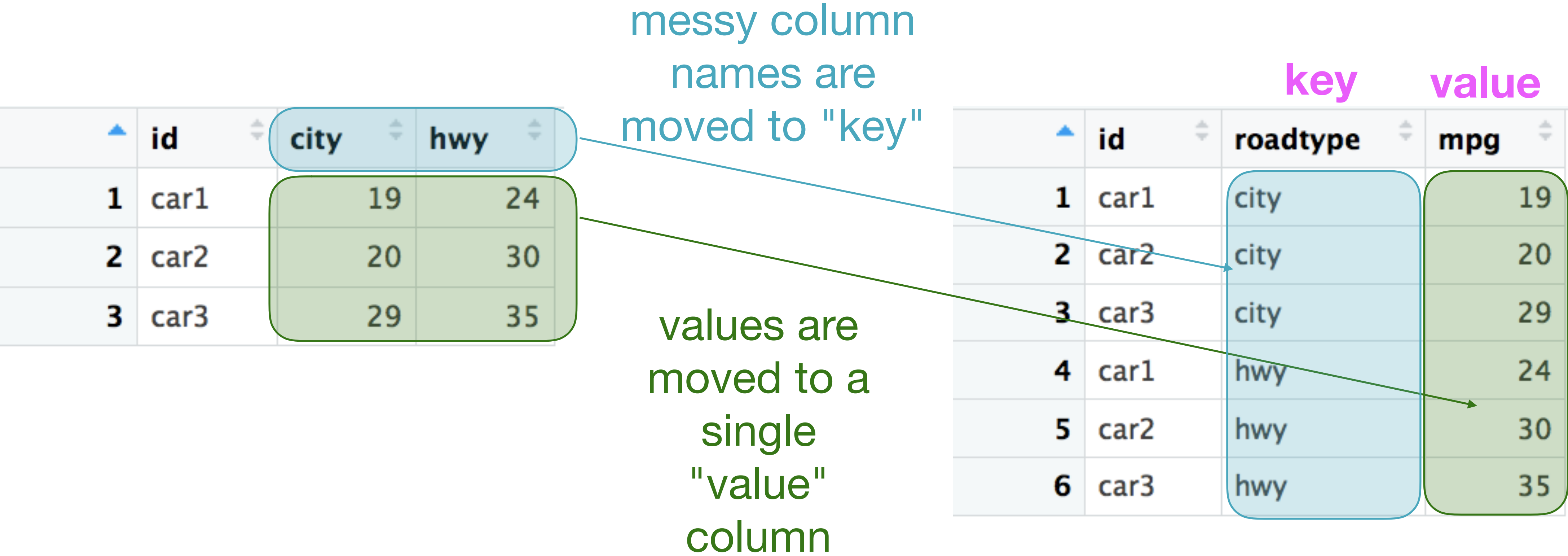
Step 1: identify "keep as is"

	id	city	hwy
1	car1	19	24
2	car2	20	30
3	car3	29	35

Step 2: create new columns

	id	roadtype	mpg
1	car1	city	19
2	car2	city	20
3	car3	city	29
4	car1	hwy	24
5	car2	hwy	30
6	car3	hwy	35

Everything else just happens:



Step 1: identify "keep as is"

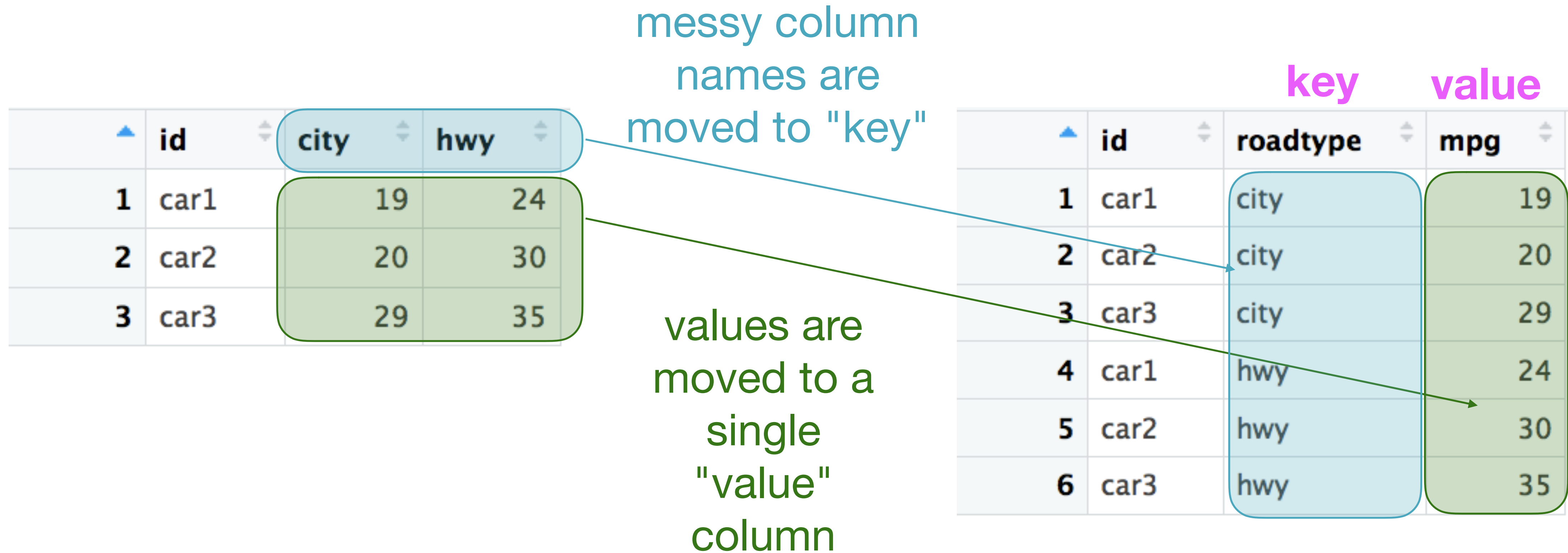
	id	city	hwy
1	car1	19	24
2	car2	20	30
3	car3	29	35

Step 2: create new columns

	id	roadtype	mpg
1	car1	city	19
2	car2	city	20
3	car3	city	29
4	car1	hwy	24
5	car2	hwy	30
6	car3	hwy	35

```
tidydata <- messydata %>%  
  gather(key = "roadtype", value = "mpg", -id)
```

Everything else just happens:



```
tidydata <- messydata %>%  
  gather(key = "roadtype", value = "mpg", -id)
```

If there is no id column

```
head(mtcars)
```

```
#>      mpg  cyl  disp  hp  drat    wt    qsec  vs  am  gear  carb
#> Mazda RX4      21.0    6  160  110  3.90  2.620  16.46  0   1     4     4
#> Mazda RX4 Wag  21.0    6  160  110  3.90  2.875  17.02  0   1     4     4
#> Datsun 710     22.8    4  108   93  3.85  2.320  18.61  1   1     4     1
#> Hornet 4 Drive  21.4    6  258  110  3.08  3.215  19.44  1   0     3     1
#> Hornet Sportabout 18.7    8  360  175  3.15  3.440  17.02  0   0     3     2
#> Valiant        18.1    6  225  105  2.76  3.460  20.22  1   0     3     1
```

```
mtcars <- mtcars %>%
  rownames_to_column("carname")
```

painters (MASS package)

labs/11-Tidying.Rmd

painters ×					
← → 📄 🔍 Filter					
	Composition	Drawing	Colour	Expression	School
Da Udine	10	8	16	3	A
Da Vinci	15	16	4	14	A
Del Piombo	8	13	16	7	A
Del Sarto	12	16	9	8	A
Fr. Penni	0	15	8	0	A
Guilio Romano	15	16	4	14	A
Michelangelo	8	17	4	8	A
Perino del Vaga	15	16	7	6	A
Perugino	4	12	10	4	A
Raphael	17	18	12	18	A
F. Zucarro	10	13	8	8	B
Fr. Salviata	13	15	8	8	B
Parmigiano	10	15	6	6	B
Primaticcio	15	14	7	10	B
T. Zucarro	13	14	10	9	B
Volterra	12	15	5	8	B
Barocci	14	15	6	10	C
Cortona	16	14	12	6	C
Showing 1 to 19 of 54 entries					