Tidy Data Article

Tiffany Cheng

February 27, 2018

Table 4 to Table 6

```r
# Read in data
raw <- readRDS("rawdat.RDS")

# Table 4 (first 10 rows only)
knitr::kable(head(raw,10))
```

religion	<$10k	$10-20k	$20-30k	$30-40k	$40-50k	$50-75k	$75-100k	$100-150k	>150k	Dor
Agnostic	27	34	60	81	76	137	122	109	84	
Atheist	12	27	37	52	35	70	73	59	74	
Buddhist	27	21	30	34	33	58	62	39	53	
Catholic	418	617	732	670	638	1116	949	792	633	
Don't know/refused	15	14	15	11	10	35	21	17	18	
Evangelical Prot	575	869	1064	982	881	1486	949	723	414	
Hindu	1	9	7	9	11	34	47	48	54	
Historically Black Prot	228	244	236	238	197	223	131	81	78	
Jehovah's Witness	20	27	24	24	21	30	15	11	6	
Jewish	19	19	25	25	30	95	69	87	151	

```r
# Tidy raw data set
raw.tidy <- raw %>% gather(key="Income", value="Frequency",-religion)
raw.order <- raw.tidy %>% arrange(religion)

# Table 6 (first 10 rows only)
knitr::kable(head(raw.order,10))
```

religion	Income	Frequency
Agnostic	<$10k	27
Agnostic	$10-20k	34
Agnostic	$20-30k	60
Agnostic	$30-40k	81
Agnostic	$40-50k	76
Agnostic	$50-75k	137
Agnostic	$75-100k	122
Agnostic	$100-150k	109
Agnostic	>150k	84
Agnostic	Don't know/refused	96

Table 7 to Table 8

```r
# Read in data
billboard <- read_csv("billboard.csv")

# Table 7 (first 8 rows)
knitr::kable(head(billboard,8))
```

year	artist.inverted	track	time	genre	date.entered	date.peaked	x1
2000	Destiny's Child	Independent Women Part I	03:38:00	Rock	2000-09-23	2000-11-18	
2000	Santana	Maria, Maria	04:18:00	Rock	2000-02-12	2000-04-08	
2000	Savage Garden	I Knew I Loved You	04:07:00	Rock	1999-10-23	2000-01-29	
2000	Madonna	Music	03:45:00	Rock	2000-08-12	2000-09-16	
2000	Aguilera, Christina	Come On Over Baby (All I Want Is You)	03:38:00	Rock	2000-08-05	2000-10-14	
2000	Janet	Doesn't Really Matter	04:17:00	Rock	2000-06-17	2000-08-26	
2000	Destiny's Child	Say My Name	04:31:00	Rock	1999-12-25	2000-03-18	
2000	Iglesias, Enrique	Be With You	03:36:00	Latin	2000-04-01	2000-06-24	

```r
# Tidy billboard data set
billboard.tidy <- billboard %>%
  gather(key="week", value="rank",-year, -artist.inverted, -track, -time, -genre, -date.entered, -date.p
  arrange(artist.inverted,track) %>%
  filter(!is.na(rank)) %>%
  separate(week,c("A","week","B"),sep=c(1,2),convert=TRUE) %>%
  separate(time,c("C","time","D"),sep=c(1,5),covert=TRUE) %>%
  select(-date.peaked,-genre,-A,-B,-C,-D) %>%
  dplyr::rename("artist"="artist.inverted","date"="date.entered") %>%
  mutate(date=date+(week-1)*7)

# Rearange column order
billboard.tidy <- billboard.tidy[c(1,2,4,3,5,6,7)]

# Table 8 (first 15 rows)
knitr::kable(head(billboard.tidy,15))
```

year	artist	time	track	date	week	rank
2000	2 Pac	4:22	Baby Don't Cry (Keep Ya Head Up II)	2000-02-26	1	87
2000	2 Pac	4:22	Baby Don't Cry (Keep Ya Head Up II)	2000-03-04	2	82
2000	2 Pac	4:22	Baby Don't Cry (Keep Ya Head Up II)	2000-03-11	3	72
2000	2 Pac	4:22	Baby Don't Cry (Keep Ya Head Up II)	2000-03-18	4	77
2000	2 Pac	4:22	Baby Don't Cry (Keep Ya Head Up II)	2000-03-25	5	87
2000	2 Pac	4:22	Baby Don't Cry (Keep Ya Head Up II)	2000-04-01	6	94
2000	2 Pac	4:22	Baby Don't Cry (Keep Ya Head Up II)	2000-04-08	7	99
2000	2Ge+her	3:15	The Hardest Part Of Breaking Up (Is Getting Back Your Stuff)	2000-09-02	1	91
2000	2Ge+her	3:15	The Hardest Part Of Breaking Up (Is Getting Back Your Stuff)	2000-09-09	2	87
2000	2Ge+her	3:15	The Hardest Part Of Breaking Up (Is Getting Back Your Stuff)	2000-09-16	3	92
2000	3 Doors Down	3:53	Kryptonite	2000-04-08	1	81
2000	3 Doors Down	3:53	Kryptonite	2000-04-15	2	70
2000	3 Doors Down	3:53	Kryptonite	2000-04-22	3	68
2000	3 Doors Down	3:53	Kryptonite	2000-04-29	4	67
2000	3 Doors Down	3:53	Kryptonite	2000-05-06	5	66