Craft-Cans Case Study

Arturo Casillas & Kevin Dickens

R Markdown Dummy

First

About the Data

Source

- The tables contain a list of 2410 US craft beers and 510 US breweries.
- The beer data corresponds to beers available in cans
- This data was traced from CraftCans.com.

About the Data

View the data

```
## Load Data ##
#getwd()
setwd("C:/Users/acasi/Downloads")
BeersData <- read.csv("Beers.csv")
BreweriesData <- read.csv("Breweries.csv")

str(BeersData)
```

```
## 'data.frame':    2410 obs. of  7 variables:
##  $ Name       : Factor w/ 2305 levels "#001 Golden Amber Lag
##  $ Beer_ID    : int  1436 2265 2264 2263 2262 2261 2260 2259
##  $ ABV        : num  0.05 0.066 0.071 0.09 0.075 0.077 0.045
##  $ IBU        : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ Brewery_id: int  409 178 178 178 178 178 178 178 178 178
##  $ Style      : Factor w/ 100 levels "","Abbey Single Ale",.
```

Munging

Basic alterations

- Column names are altered for clarity and to minimize merging issues
- The beers and breweries are linked by a numeric ID, which is used to merge the two tables
- 'Brew_ID'

```
## Prepare for merging ##
names(BeersData)
```

```
## [1] "Name"        "Beer_ID"    "ABV"          "IBU"          "Br
## [6] "Style"       "Ounces"
```

```
names(BeersData)[5] = "Brew_ID"
names(BeersData)[1] = "Beer.name"
names(BreweriesData)[2] = "Brewery.name"
```

Munging

View the final data

-first 6 rows

```
library(knitr)
#library(kableExtra)
kable(head(AllBeer, 6))
```

Brew_ID	Beer.name	Beer_ID	ABV	IBU	Style
1	Get Together	2692	0.045	50	American IPA
1	Maggie's Leap	2691	0.049	26	Milk / Sweet Stout
1	Wall's End	2690	0.048	19	English Brown Ale
1	Pumpion	2689	0.060	38	Pumpkin Ale
1	Stronghold	2688	0.060	25	American Porter
1	Parapet ESB	2687	0.056	47	Extra Special / Strong

Munging

View the final data

-Try an xtable

```
print(xtable(head(AllBeer)), comment=F, auto = TRUE)
```

	Brew_ID	Beer.name	Beer_ID	ABV	IBU	Style
1	1	Get Together	2692	0.04	50	American IPA
2	1	Maggie's Leap	2691	0.05	26	Milk / Sweet Stou
3	1	Wall's End	2690	0.05	19	English Brown Ale
4	1	Pumpion	2689	0.06	38	Pumpkin Ale
5	1	Stronghold	2688	0.06	25	American Porter
6	1	Parapet ESB	2687	0.06	47	Extra Special / St

Munging

View the final data

last 6 rows

```
library(knitr)
kable(tail(AllBeer, 6))
```

	Brew_ID	Beer.name	Beer_ID	ABV	IBU	Sty
2405	556	Pilsner Ukiah	98	0.055	NA	Ger
2406	557	Heinnieweisse Weissebier	52	0.049	NA	Hef
2407	557	Snapperhead IPA	51	0.068	NA	Am
2408	557	Moo Thunder Stout	50	0.049	NA	Mil
2409	557	Porkslap Pale Ale	49	0.043	NA	Am
2410	558	Urban Wilderness Pale Ale	30	0.049	NA	Eng

Data Integrity

Missing Vlaues

- Check the missing values number and indicators

```
## Check for missing values ##
apply(apply(AllBeer, 2, is.na), 2, sum)
```

```
##      Brew_ID    Beer.name      Beer_ID           ABV
##            0            0            0           62
##        Style       Ounces Brewery.name          City
##            0            0            0            0
```

```
t=sapply(AllBeer, function(y) sum(length(which(is.na(y)))))
summary(t)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
```

Data Integrity

Notes and Observations

- Only ABV and IBU have missing values
- Almost 50% of IBU values are missing -This certainly affect any analysis that hinges on IBU values
- Spot checks on the data suggest that character variables use blanks for missing values
- This needs to be verified

Data Integrity

Double Check Missing Vlaues

- Check for blanks instead of NA's

```
## Double Check ##
## Look at balnk Strings ##
apply(AllBeer, 2, function(y) sum(y == ""))
```

```
##       Brew_ID     Beer.name       Beer_ID            ABV
##             0             0             0             NA
##         Style        Ounces Brewery.name          City
##             5             0             0             0
```

- 'Stye', which represents style of beer, has 5 additional missing values

External Data

- From R
- R contains additional region and division data per state in the package 'Datasets'
- This data is from a 1977 report from the chamber of commerce

External Data

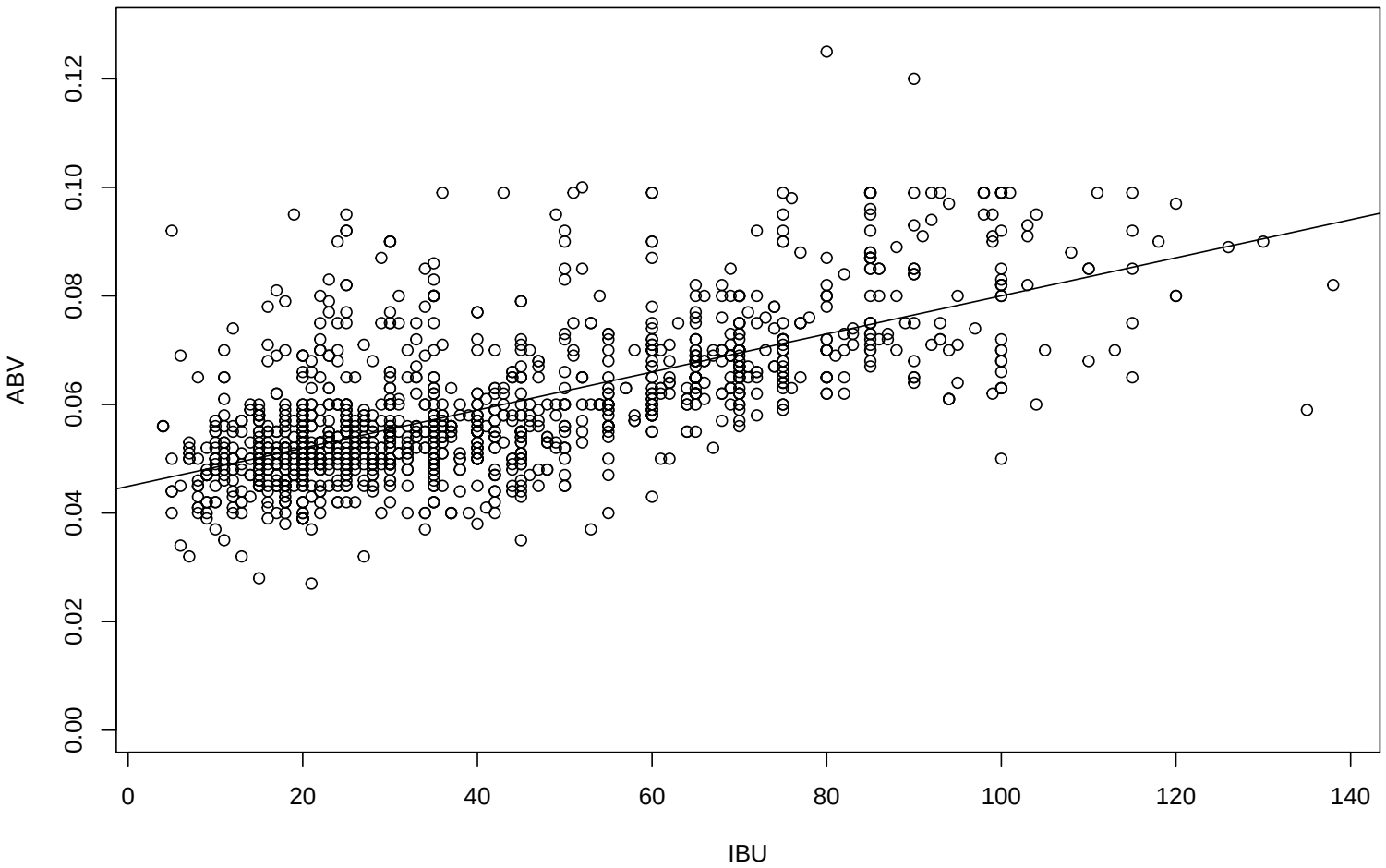
- From census
- This is a 2016 geography table

Tail2

-experiment with plotting last 6 rows

Brew_ID	Beer.name	Beer_ID	ABV	IBU	Style
1	Get Together	2692	0.045	50	American IPA
1	Maggie's Leap	2691	0.049	26	Milk / Sweet Stout
1	Wall's End	2690	0.048	19	English Brown Ale
1	Pumpion	2689	0.060	38	Pumpkin Ale
1	Stronghold	2688	0.060	25	American Porter
1	Parapet ESB	2687	0.056	47	Extra Special / Strong

Simple Plot



Fancier Plot