

Wine Retailer AB Test

When your sample size is large

true

true

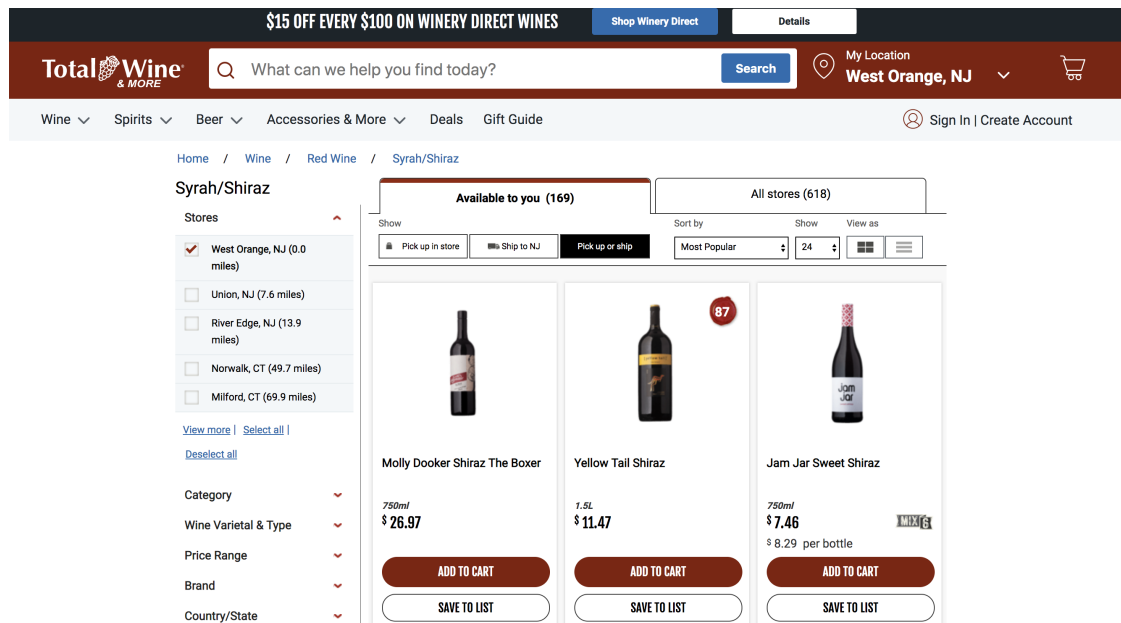
Spring 2020

“When customers are randomly assigned to treatment and control groups, and there are many customers in each group, then you may effectively have multiple experiments to analyze.”

- Anderson and Simester (2011) A step-by-step guide to smart business experiments, *HBR*

Other times, we actually have multiple experiments! Same basic idea of analysis approach.

Wine retailer experiment



Wine retailer experiment

Test setting: email to retailer customers

Unit: customer (email address)

Treatments: email version A, email version B, holdout

Reponse: open, click and 1-month purchase (\$)

Selection: all active customers

Assignment: randomly assigned (1/3 each) to email A, email B or control (ctrl)

```
summary(d)
```

```
##      user_id      cpgn_id      group      open
## Min.   :1000001  1901Email:123988  ctrl   :41330  Min.   :0.0000
## 1st Qu.:1030998                email_A:41329  1st Qu.:0.0000
## Median :1061994                email_B:41329  Median :0.0000
## Mean   :1061994                                Mean   :0.4569
## 3rd Qu.:1092991                                3rd Qu.:1.0000
## Max.   :1123988                                Max.   :1.0000
##      click      purch      chard      sav_blanc
## Min.   :0.00000  Min.   : 0.00  Min.   : 0.00  Min.   : 0.00
## 1st Qu.:0.00000  1st Qu.: 0.00  1st Qu.: 0.00  1st Qu.: 0.00
## Median :0.00000  Median : 0.00  Median : 0.00  Median : 0.00
## Mean   :0.07374  Mean   : 13.27  Mean   : 73.31  Mean   : 72.45
## 3rd Qu.:0.00000  3rd Qu.: 0.00  3rd Qu.: 54.06  3rd Qu.: 57.42
## Max.   :1.00000  Max.   :1820.70  Max.   :9636.92  Max.   :6609.92
##      syrah      cab      past_purch      last_purch
## Min.   : 0.00  Min.   : 0.00  Min.   : 0.00  Min.   : 0.00
## 1st Qu.: 0.00  1st Qu.: 0.00  1st Qu.: 0.00  1st Qu.: 26.00
## Median : 0.00  Median : 0.00  Median : 91.22  Median : 63.00
## Mean   : 26.68  Mean   : 16.35  Mean   : 188.79  Mean   : 89.98
## 3rd Qu.: 20.91  3rd Qu.: 12.96  3rd Qu.: 246.87  3rd Qu.:125.00
## Max.   :2880.15  Max.   :2365.90  Max.   :9636.92  Max.   :992.00
##      visits
## Min.   : 0.000
## 1st Qu.: 4.000
## Median : 6.000
## Mean   : 5.946
## 3rd Qu.: 7.000
## Max.   :51.000
```

Types of variables associated with a test

- **Treatment indicator** (W's)
 - Which (randomized) treatment was received
 - here these are group (email A, B, or ctrl)
- **Response** (Y's)
 - Outcome(s) measured for each customer. Aka the DV or dependant variable.
 - Here these are open, click, and purch
- **Baseline variables** (“X's”)
 - Other stuff we know about customers **prior** to the randomization
 - here includes
 - amount purchased of specific wine types in past (chard, sav_blanc, syrah, cab)
 - total past purchases, sum of chard+sav_blanc+syrah+cab in past year (past_purch)
 - days since last purchase or 2 years if more than 2 years ago (last_purch)
 - number of website visits (visits)

We can depict the average effects in table and graph form as follows:

```
dt = data.table(d)
dagg = dt[,.(open = mean(open),
             click=mean(click),
```

```

    purch = mean(purch),
    seOpen = sd(open)/sqrt(.N),
    seClick=sd(click)/sqrt(.N),
    sePurch = sd(purch)/sqrt(.N),.N),
  by = .(group)]
dagg

```

```

##      group      open      click      purch      seOpen      seClick      sePurch
## 1:   ctrl 0.0000000 0.00000000 12.32166 0.000000000 0.000000000 0.2078149
## 2: email_B 0.6536814 0.08969489 13.68348 0.002340447 0.001405579 0.2211614
## 3: email_A 0.7170994 0.13152992 13.81592 0.002215565 0.001662522 0.2304795
##      N
## 1: 41330
## 2: 41329
## 3: 41329

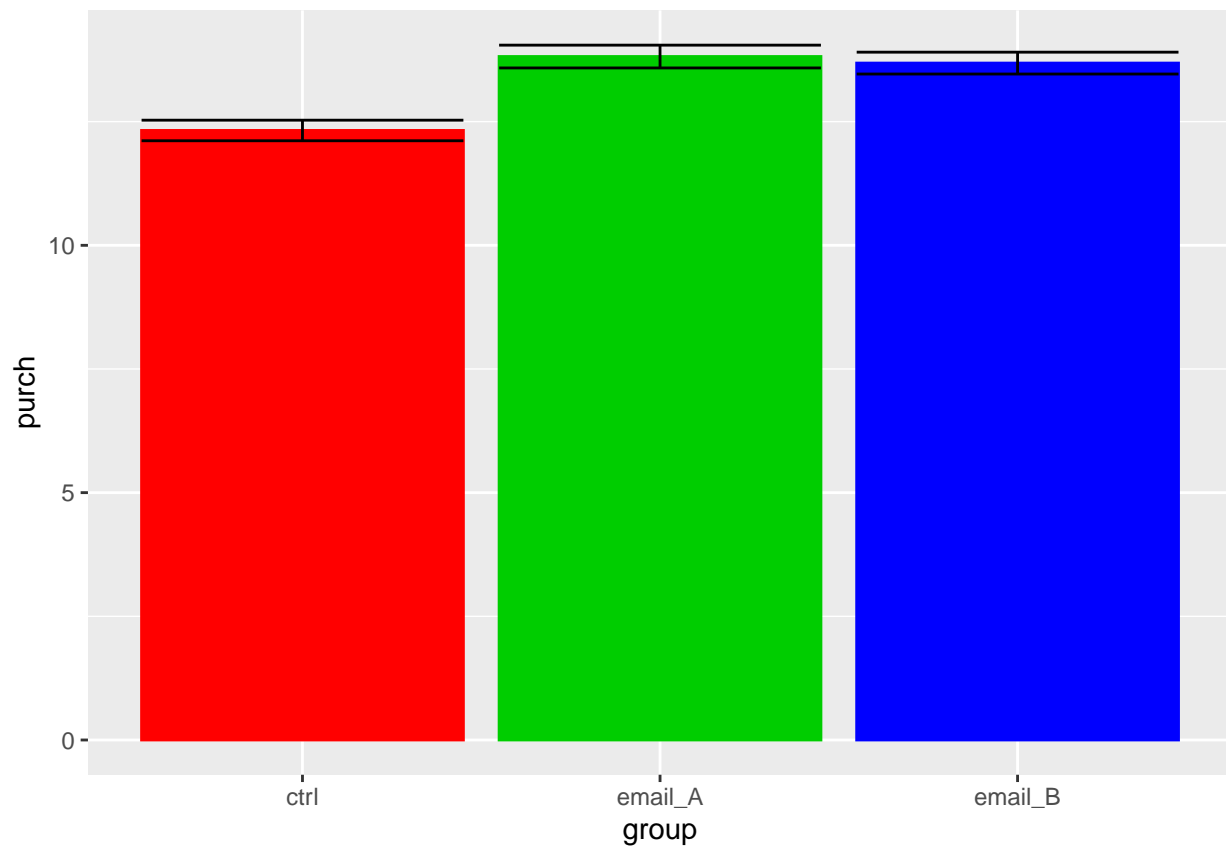
```

For this case explanation, we will focus on the purchases variable.

```

dodge = position_dodge(width=1); ##to form constant dimensions
ggplot(aes(x=group,
  y=purch,
  ymax=purch+sePurch,
  ymin=purch-sePurch),
  data=dagg)+
  geom_bar(position=dodge,stat="identity",col=2:4,fill=2:4) +
  geom_errorbar(position=dodge)

```



```
labs(x="Group",y="Purchases")
```

```
## $x
## [1] "Group"
##
## $y
## [1] "Purchases"
##
## attr("class")
## [1] "labels"
```

To evaluate the average causal effect on purchases we use lm as follows.

```
summary(lm(purch~group,data=d)) #compares each email to control group
```

```
##
## Call:
## lm(formula = purch ~ group, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.82  -13.82  -13.68  -12.32  1806.88
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   12.3217     0.2200   56.004 < 2e-16 ***
## groupemail_A    1.4943     0.3111    4.802 1.57e-06 ***
## groupemail_B    1.3618     0.3111    4.377 1.21e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.73 on 123985 degrees of freedom
## Multiple R-squared:  0.0002279, Adjusted R-squared:  0.0002118
## F-statistic: 14.13 on 2 and 123985 DF, p-value: 7.288e-07
```

All effects are both highly significant and the effect sizes are \$1.49 for email A and \$1.36 for email B. Notice further that the two emails do not appear to be significantly different from one another. As a result, we create a variable, email, that pools the two together. We then rerun the analysis with the pooled variables.

```
d$email = (d$group != "ctrl")
names(d)
```

```
## [1] "user_id"      "cpgn_id"      "group"        "open"         "click"
## [6] "purch"        "chard"        "sav_blanc"    "syrah"        "cab"
## [11] "past_purch"   "last_purch"   "visits"       "email"
```

```
summary(lm(purch~email,data=d)) #combining email groups
```

```
##
## Call:
## lm(formula = purch ~ email, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.75  -13.75  -13.75  -12.32  1806.95
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.3217      0.2200   56.0 < 2e-16 ***
## emailTRUE   1.4280      0.2695    5.3 1.16e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.73 on 123986 degrees of freedom
## Multiple R-squared:  0.0002265, Adjusted R-squared:  0.0002184
## F-statistic: 28.09 on 1 and 123986 DF, p-value: 1.162e-07
```

Notice that the standard error goes down because of the larger sample size. The results are also easier to interpret. Often we combine treatment groups in this way for ease of communication and for better understanding how precise our estimates are.

The precision of our estimates is sufficient here to establish the average effect, but if they didn't we could try to add covariates to absorb some of the error and reduce the standard errors. We can do this as follows here:

```
#adding baseline variables as controls
summary(lm(purch~email+chard+sav_blanc+syrah+cab+past_purch+last_purch+visits,data=d))
```

```
##
## Call:
## lm(formula = purch ~ email + chard + sav_blanc + syrah + cab +
##     past_purch + last_purch + visits, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -247.78  -14.65  -10.23   -1.58  1784.98
##
## Coefficients: (1 not defined because of singularities)
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.6501221  0.3658180  37.314 < 2e-16 ***
## emailTRUE    1.4036290  0.2618889   5.360 8.35e-08 ***
## chard         0.0262443  0.0006489  40.442 < 2e-16 ***
## sav_blanc     0.0273778  0.0006708  40.813 < 2e-16 ***
## syrah         0.0303311  0.0016749  18.109 < 2e-16 ***
## cab          0.0238069  0.0026511   8.980 < 2e-16 ***
## past_purch      NA         NA      NA      NA
## last_purch    -0.0737448  0.0013748 -53.639 < 2e-16 ***
## visits        0.0366075  0.0508939   0.719  0.472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.47 on 123980 degrees of freedom
## Multiple R-squared:  0.05567, Adjusted R-squared:  0.05562
## F-statistic: 1044 on 7 and 123980 DF, p-value: < 2.2e-16
```

```
summary(lm(purch~email+chard+sav_blanc+syrah+cab+last_purch+visits,data=d))
```

```
##
## Call:
## lm(formula = purch ~ email + chard + sav_blanc + syrah + cab +
##     last_purch + visits, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -247.78 -14.65 -10.23 -1.58 1784.98
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.6501221  0.3658180  37.314 < 2e-16 ***
## emailTRUE   1.4036290  0.2618889   5.360 8.35e-08 ***
## chard        0.0262443  0.0006489  40.442 < 2e-16 ***
## sav_blanc    0.0273778  0.0006708  40.813 < 2e-16 ***
## syrah        0.0303311  0.0016749  18.109 < 2e-16 ***
## cab          0.0238069  0.0026511   8.980 < 2e-16 ***
## last_purch  -0.0737448  0.0013748 -53.639 < 2e-16 ***
## visits       0.0366075  0.0508939   0.719  0.472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.47 on 123980 degrees of freedom
## Multiple R-squared:  0.05567,    Adjusted R-squared:  0.05562
## F-statistic: 1044 on 7 and 123980 DF,  p-value: < 2.2e-16
```

#adding controls separating emails

```
summary(lm(purch~group+chard+sav_blanc+syrah+cab+last_purch+visits,data=d))
```

```
##
## Call:
## lm(formula = purch ~ group + chard + sav_blanc + syrah + cab +
##     last_purch + visits, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -247.77  -14.65  -10.23   -1.58  1784.94
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.6502257  0.3658195  37.314 < 2e-16 ***
## groupemail_A  1.4489604  0.3024059   4.791 1.66e-06 ***
## groupemail_B  1.3582981  0.3024053   4.492 7.07e-06 ***
## chard         0.0262439  0.0006489  40.441 < 2e-16 ***
## sav_blanc     0.0273772  0.0006708  40.812 < 2e-16 ***
## syrah         0.0303314  0.0016749  18.110 < 2e-16 ***
## cab           0.0238054  0.0026511   8.979 < 2e-16 ***
## last_purch   -0.0737456  0.0013749 -53.639 < 2e-16 ***
## visits        0.0366162  0.0508941   0.719  0.472
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43.47 on 123979 degrees of freedom
## Multiple R-squared:  0.05567,    Adjusted R-squared:  0.05561
## F-statistic: 913.6 on 8 and 123979 DF,  p-value: < 2.2e-16
```

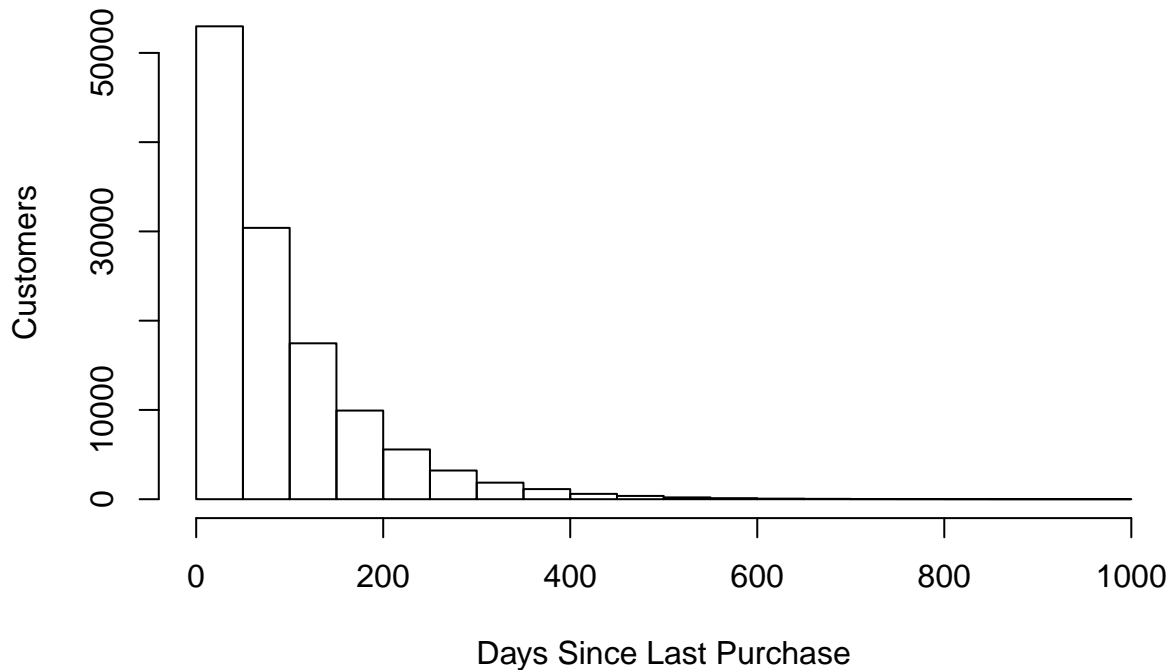
First, the first model included all of the purchases by product category—`chard`, `sav_blanc`, `syrah`, and `cab`—along with the total past purchases, `past_purch`. These are perfectly collinear because `past_purch=chard+sav_blanc+syrah+cab`. Hence, in the first regression above, the row for `past_purch` is NA. In the following regressions, we drop the `past_purch` variable.

Notice the standard errors do shrink some, but not by much. We now plot the purchase rates for `last_purch` and then create a dummy variable for `last_purch` to capture new versus

older customers.

Baseline variable: days since last purchase

Histogram of Days Since Last Purchase



Experiments within experiments

Consider the customers who have made a purchase in the last 60 days.

Within that subset, customers were randomly assigned to receive email A, email B or no email.

So, we can analyze the data for a subgroup as its own test by slicing down and then re-analyzing.

However, we will only have enough sample in the subgroup, if our test is big enough.

Slicing and dicing: recent buyers versus aged customers

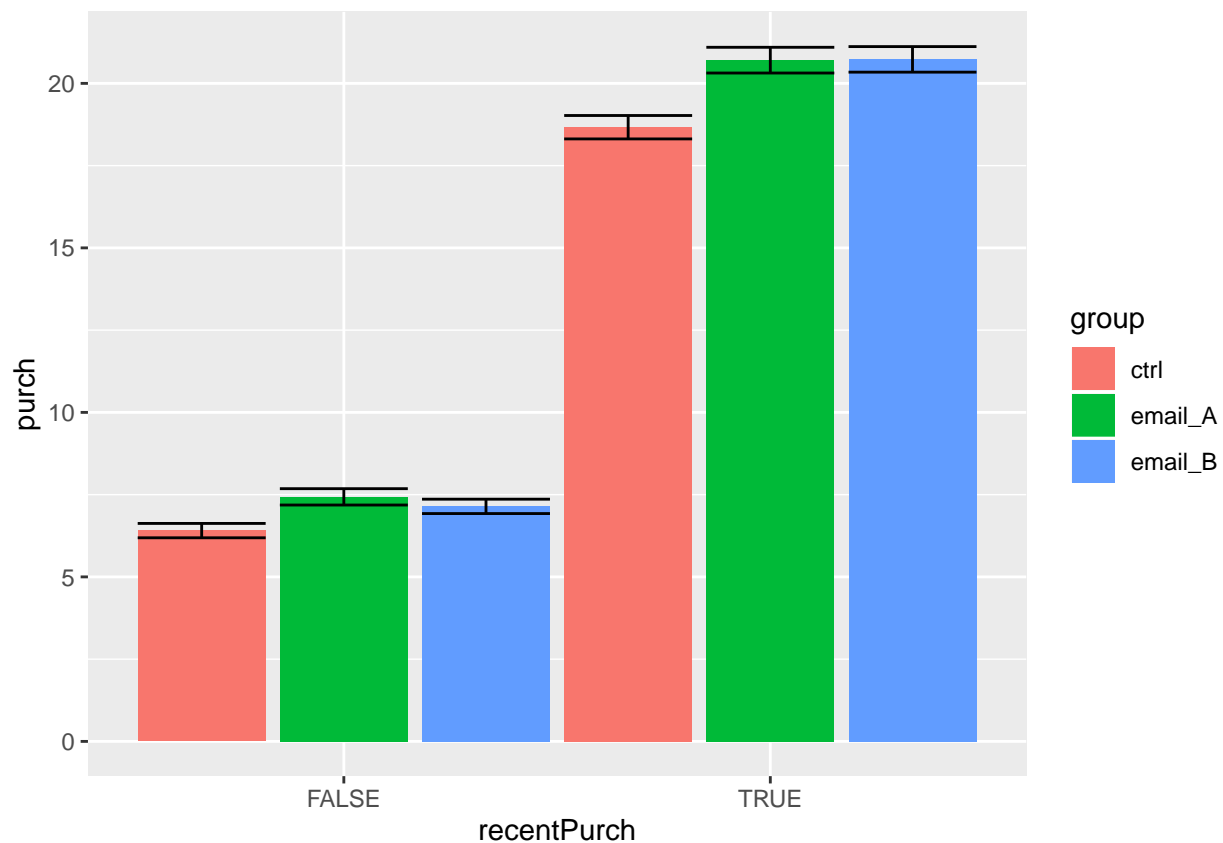
```
dagg = dt[,.(open = mean(open),
             click=mean(click),
             purch = mean(purch),
             seOpen = sd(open)/sqrt(.N),
             seClick=sd(click)/sqrt(.N),
             sePurch = sd(purch)/sqrt(.N),.N),
          by = .(group,recentPurch)]
dagg
```

##	group	recentPurch	open	click	purch	seOpen	seClick
## 1:	ctrl	FALSE	0.0000000	0.00000000	6.406703	0.000000000	0.000000000
## 2:	email_B	FALSE	0.5072782	0.07068209	7.143443	0.003414936	0.001750634
## 3:	email_A	TRUE	0.8656671	0.16114912	20.705377	0.002418873	0.002607971
## 4:	email_A	FALSE	0.5794528	0.10408801	7.432902	0.003370408	0.002084967

```
## 5: email_B      TRUE 0.8114099 0.11017844 20.729422 0.002773438 0.002219927
## 6:   ctrl      TRUE 0.0000000 0.00000000 18.667363 0.000000000 0.000000000
##   sePurch      N
## 1: 0.2182475 21391
## 2: 0.2186549 21434
## 3: 0.3913648 19876
## 4: 0.2484566 21453
## 5: 0.3883006 19895
## 6: 0.3561425 19939
```

- Recent buyers buy more on average
- The email seems to produce a stronger effect on purchases for more recent buyers (~\$2 versus \$1)

Is email more effective for recent buyers?



```
## $x
## [1] "Group"
##
## $y
## [1] "Purchases"
##
## attr("class")
## [1] "labels"
```


We slice based on baseline variables

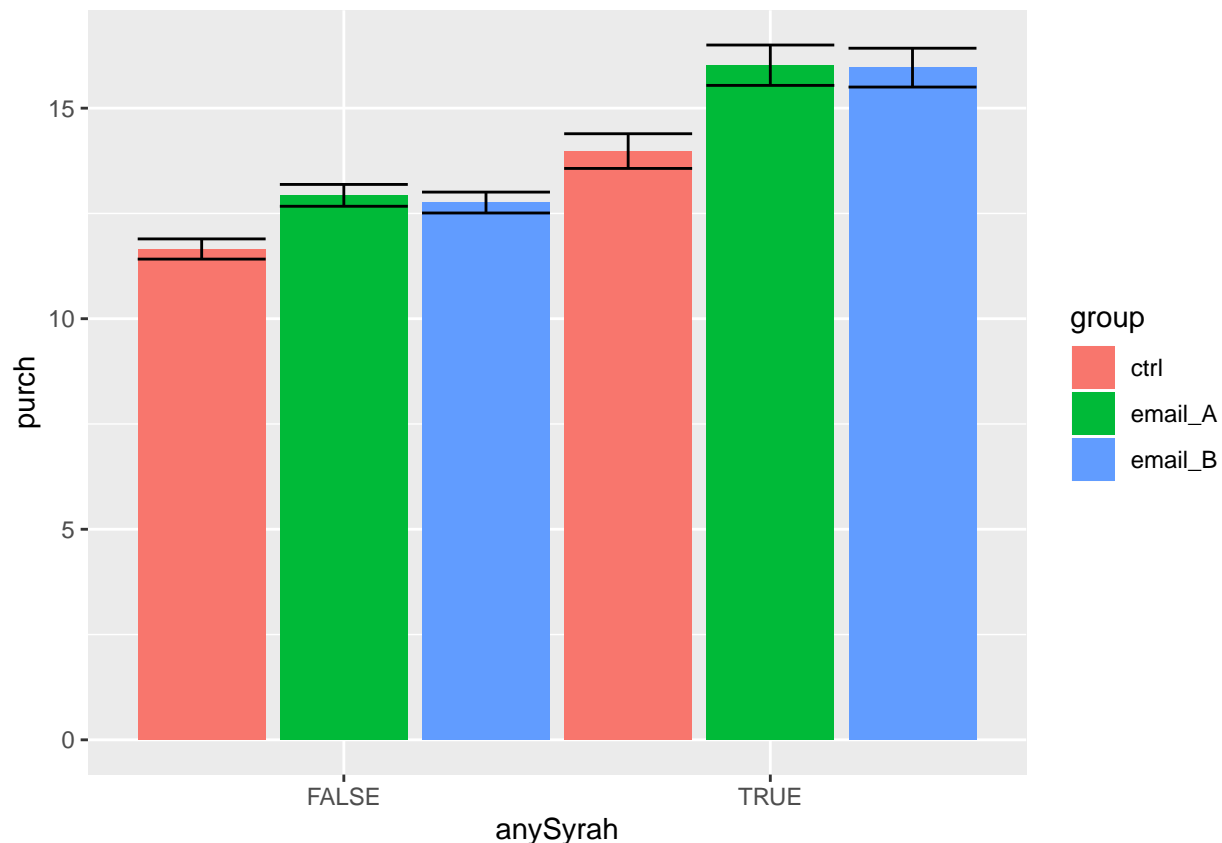
Anyone who keeps historic data on customers or visitors has lots of baseline variables available for slicing and dicing:

- website visits (to particular parts of the site)
- sign-ups
- geographic location
- source
- past purchase (by category)
- recency
- frequency

Exercise

Re-analyze the opens, clicks and purchases for people who have bought syrah in the past.

```
d$anySyrah = (d$syrah > 0); dt = data.table(d);
dagg = dt[,.(open = mean(open),
              click=mean(click),
              purch = mean(purch),
              seOpen = sd(open)/sqrt(.N),
              seClick=sd(click)/sqrt(.N),
              sePurch = sd(purch)/sqrt(.N),.N),
          by = .(group,anySyrah)]
ggplot(aes(fill=group,
           y=purch,
           x=anySyrah,
           ymax=purch+sePurch,
           ymin=purch-sePurch),
       data=dagg)+
geom_bar(position=dodge,stat="identity") +
geom_errorbar(position=dodge)
```



```
labs(x="Group",y="Purchases")
```

```
## $x
## [1] "Group"
##
## $y
## [1] "Purchases"
##
## attr(,"class")
## [1] "labels"
```

Repeated significance testing

Slicing and dicing means you will run many significance tests. This means repeated hypothesis testing. Such repeated testing can lead to false positives. Remember every 20 variables you test at 95% significance roughly would lead to one being significant by chance! For experiments, this means you should consider several possibilities.

First, you could view any single test as less significant than the test states. There are formal corrections for doing this such as the Bonferroni adjustments to p-values. These corrections are generally conservative and don't necessarily provide the desired guidance on which actions are optimal to take, but rather on which actions are significantly different. This is not the right decision criterion!

Second, if you think you've found a targeting criterion, you could consider re-testing the effect. Of course, you should consider value of information, which in this case you can calculate exactly if you want, since you can evaluate conservatively (using Bonferroni adjustments) how uncertain you are as well as how certain you'd be if you ran another experiment.

Next we will discuss constructing conditional causal effects using regression analysis. We first we consider

just the `recentPurch` variable as a conditioning variable and pooling over the emails.

Measuring causal effects with regression: Conditional causal effects

```
#compares each email to control group
lmInteraction = lm(purch~email*recentPurch,data=d)
summary(lmInteraction)

##
## Call:
## lm(formula = purch ~ email * recentPurch, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.72  -18.67   -7.29   -6.41  1799.98
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.4067    0.3026   21.18  <2e-16 ***
## emailTRUE        0.8815    0.3704    2.38  0.0173 *
## recentPurchTRUE  12.2607    0.4356   28.15  <2e-16 ***
## emailTRUE:recentPurchTRUE  1.1685    0.5335    2.19  0.0285 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.25 on 123984 degrees of freedom
## Multiple R-squared:  0.02148,    Adjusted R-squared:  0.02145
## F-statistic: 907.1 on 3 and 123984 DF,  p-value: < 2.2e-16
```

The main effect of the email variable is significant still, leading to \$0.88 more sales. However, the interpretation is different. Now this effect relates to the omitted group from the interaction effect (when the `recentPurch` variable=0 or FALSE). Hence, the effect of `emailTRUE` is for those that have not bought recently. This group of customers are still significantly affected by the email.

This analysis reveals two other things about subgroups of customers.

1. Subgroups will vary in how much they engage in behaviors (main effect of baseline variables)
 - Recent buyers tend to have \$12.26 higher average purchases in the future
2. Subgroups vary in how much they respond to treatments (interaction effects)
 - Recent buyers are more affected by the email, leading to addition \$1.17 in spending

Since the segmentation variable is categorical, to interpret the exact level of effect for each of the subgroups we use the main for the omitted group, and we add the main and interaction effects together for the included group. Here that means that the customers without recent purchases have an effect of \$0.88 and the customers with recent purchase have an effect of \$2.05.

We can also obtain this by rerunning the analysis in a slightly different formula as shown below:

```
lmCellMeans = lm(purch~recentPurch + email:recentPurch,data=d)
summary(lmCellMeans) #compares each email to control group

##
## Call:
## lm(formula = purch ~ recentPurch + email:recentPurch, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -20.72 -18.67 -7.29 -6.41 1799.98
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.4067 0.3026 21.175 < 2e-16 ***
## recentPurchTRUE 12.2607 0.4356 28.147 < 2e-16 ***
## recentPurchFALSE:emailTRUE 0.8815 0.3704 2.380 0.0173 *
## recentPurchTRUE:emailTRUE 2.0500 0.3840 5.339 9.36e-08 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.25 on 123984 degrees of freedom
## Multiple R-squared: 0.02148, Adjusted R-squared: 0.02145
## F-statistic: 907.1 on 3 and 123984 DF, p-value: < 2.2e-16
stargazer(lmInteraction,lmCellMeans,type="text")
```

```
##
## =====
## Dependent variable:
## -----
##           purch
##           (1)      (2)
## -----
## email           0.882**
##                (0.370)
##
## recentPurch     12.261***    12.261***
##                (0.436)      (0.436)
##
## emailTRUE:recentPurch 1.169**
##                (0.534)
##
## recentPurchFALSE:email           0.882**
##                                (0.370)
##
## recentPurchTRUE:email           2.050***
##                                (0.384)
##
## Constant        6.407***    6.407***
##                (0.303)      (0.303)
## -----
## Observations    123,988    123,988
## R2               0.021      0.021
## Adjusted R2     0.021      0.021
## Residual Std. Error (df = 123984) 44.250    44.250
## F Statistic (df = 3; 123984) 907.150***    907.150***
## =====
## Note:           *p<0.1; **p<0.05; ***p<0.01
```

When considering subgroups, it is sometimes helpful to also look at the individual treatments again. The reason is that some subgroups might respond more to particular treatments than others.

```
#compares each email to control group
summary(lm(purch~group*recentPurch,data=d))
```

```
##
## Call:
## lm(formula = purch ~ group * recentPurch, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.73  -18.67   -7.14   -6.41  1799.99
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      6.4067    0.3026  21.175  <2e-16 ***
## groupemail_A      1.0262    0.4276   2.400   0.0164 *
## groupemail_B      0.7367    0.4277   1.723   0.0849 .
## recentPurchTRUE   12.2607    0.4356  28.147  <2e-16 ***
## groupemail_A:recentPurchTRUE  1.0118    0.6161   1.642   0.1005
## groupemail_B:recentPurchTRUE  1.3253    0.6161   2.151   0.0315 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.25 on 123982 degrees of freedom
## Multiple R-squared:  0.02148,    Adjusted R-squared:  0.02144
## F-statistic: 544.4 on 5 and 123982 DF,  p-value: < 2.2e-16
```

We still see the baseline variable has a significant positive effect on sales, indicating that purchases are higher for those that bought recently. In terms of the causal effects, both emails have a (marginally) significant main effect. Again, these main effects can be interpreted as for those not buying within the last 60 days. Email B also has a significant interaction effect. This interaction effect for email B means that those buying recently have greater purchases than those that did not buy recently. Email A is almost marginally significant (at p-value of 0.1005). Interpreting this strictly means that for email A those buying recently were not affected differently from those not buying recently.

To understand the effect of the email on each group we need to add the main effect and interaction effects for the same email together. Instead of doing all of the math in our heads, we rerun the analysis.

```
#compares each email to control group
summary(lm(purch~0+recentPurch + group:recentPurch,data=d))
```

```
##
## Call:
## lm(formula = purch ~ 0 + recentPurch + group:recentPurch, data = d)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.73  -18.67   -7.14   -6.41  1799.99
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## recentPurchFALSE      6.4067    0.3026  21.175  < 2e-16 ***
## recentPurchTRUE     18.6674    0.3134  59.568  < 2e-16 ***
## recentPurchFALSE:groupemail_A  1.0262    0.4276   2.400   0.0164 *
## recentPurchTRUE:groupemail_A   2.0380    0.4435   4.595 4.33e-06 ***
## recentPurchFALSE:groupemail_B  0.7367    0.4277   1.723   0.0849 .
## recentPurchTRUE:groupemail_B   2.0621    0.4434   4.650 3.32e-06 ***
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 44.25 on 123982 degrees of freedom
## Multiple R-squared:  0.1007, Adjusted R-squared:  0.1006
## F-statistic: 2313 on 6 and 123982 DF,  p-value: < 2.2e-16
```

Here we can see that the effect for both emails is almost the same among those with recent purchases, but that email B is 25% lower for those without recent purchases. We can guess based on the value of the standard errors that the difference between emails is not significant. We can test this explicitly a number of different ways. Here we do so simply by dropping the control group. Note that strictly we are losing power here, so this is not the tightest way to do the test.

```
#compares across emails
```

```
summary(lm(purch~group*recentPurch,data=d,subset=d$email))
```

```
##
## Call:
## lm(formula = purch ~ group * recentPurch, data = d, subset = d$email)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20.73  -20.71   -7.43   -7.14  1799.99
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      7.4329     0.3101  23.966 <2e-16 ***
## groupemail_B     -0.2895     0.4387  -0.660  0.509
## recentPurchTRUE    13.2725     0.4472  29.678 <2e-16 ***
## groupemail_B:recentPurchTRUE  0.3135     0.6325   0.496  0.620
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 45.43 on 82654 degrees of freedom
## Multiple R-squared:  0.02136, Adjusted R-squared:  0.02132
## F-statistic: 601.3 on 3 and 82654 DF,  p-value: < 2.2e-16
```

In this model set-up the omitted category for group is email A, not the control group. So the intercept represents average purchases for email A recipients without recent purchases. Likewise, the `recentPurchaseTRUE` represents the average purchases for email A recipients with recent purchases. The email B effects are in contrast to email A. Interpreting these we see that for those without recent purchases, email B is not significantly different than email A (-0.29) and for those with recent purchases, email B is also not significantly different from email A (0.31). Thus, we are probably safe in pooling over the emails for this analysis.

After we identify an uplift model we are confident in, we can use it to identify targets. While such targets can sometimes be identified directly by looking at the coefficients, if many variables are present, it can be helpful to use the conditional means to create scores for each potential target. In this case, these represent the treatment effect for each group. This takes us back to selecting a model of the treatment effects we are confident about. Assume, we decide to select the model with only recent purchases as the breakouts. Then we use the following model, which controls for recentPurchase while also allowing for distinct effects for those with and without a recent purchase.

```
lmToUse = lm(purch~recentPurch + email:recentPurch,data=d)
stargazer(lmToUse,type="text")
```

```
##
## =====
```

```

##                               Dependent variable:
##                               -----
##                               purch
## -----
## recentPurch                  12.261***
##                               (0.436)
##
## recentPurchFALSE:email       0.882**
##                               (0.370)
##
## recentPurchTRUE:email        2.050***
##                               (0.384)
##
## Constant                     6.407***
##                               (0.303)
## -----
## Observations                 123,988
## R2                           0.021
## Adjusted R2                  0.021
## Residual Std. Error         44.250 (df = 123984)
## F Statistic                  907.150*** (df = 3; 123984)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01

```

We can then compare the effects against a threshold for the treatment effect size. For instance, if we know that we have a margin of 25% and the opportunity cost for an email is 30 cents, then the formula is

$$S_i * M_i - C_i > 0 \quad (1)$$

where S_i is the score for case i , M_i is the margin, and C_i is the cost. Hence, the score must be bigger than the cost divided by the margin, or in this example, 120 cents. Our estimated coefficients are 0.882 and 2.050. Hence, we would only want to send an email to the people who had a recent purchase.

##Some further ideas NOT in the wine retailer

Notice that in this case, with or without the email consumers can choose to purchase wine from the retailer. Hence, this is considered a *lift* action. If consumers could only choose to purchase if they received an offer, it would be referred to as an *exclusive* action. Exclusive actions allow for simpler models because we don't need a control group. The control group is by definition no response (i.e., 0)! If we have a single treatment, this means we can directly model the treatment effect as main effects of the baseline variables. With multiple treatments, we need to interact the treatments with those variables, still.

Because when we have exclusive actions the comparison is zero, we can use the `predict()` function to create our scores. This greatly simplifies evaluating who to target. With such exclusive actions, we just score them and compare against the threshold.

Things you just learned

- Large sample -> look for heterogeneous treatment effects using baseline variables
- Two ways to find heterogeneous treatment effects
 - Slicing and dicing - just getting a sense
 - Uplift modeling - measuring the effect and whether it is significant
- How to target based on an uplift model
- The difference between randomized control trials with lift models vs. exclusive actions
 - For estimating effects
 - For creating scores to use for targeting decisions