Datacamp_Importing & Cleaning Data in R: Case Studies_Ticket Sales Data

dizhen

2019/4/5

Importing the data

```r
# Import sales.csv: sales
sales <- read.csv("data/sales.csv",stringsAsFactors = FALSE)
```

```r
# View dimensions of sales
dim(sales)
```

```
## [1] 2628   46
```

```r
# Inspect first 6 rows of sales
head(sales,n = 6)
```

```
##   X            event_id       primary_act_id     secondary_act_id
## 1 1 abcaf1adb99a935fc661 43f0436b905bfa7c2eec b85143bf51323b72e53c
## 2 2 6c56d7f08c95f2aa453c 1a3e9aecd0617706a794 f53529c5679ea6ca5a48
## 3 3 c7ab4524a121f9d687d2 4b677c3f5bec71eec8d1 b85143bf51323b72e53c
## 4 4 394cb493f893be9b9ed1 b1ccea01ad6ef8522796 b85143bf51323b72e53c
## 5 5 55b5f67e618557929f48 91c03a34b562436efa3c b85143bf51323b72e53c
## 6 6 4f10fd8b9f550352bd56 ac4b847b3fde66f2117e 63814f3d63317f1b56c4
##     purch_party_lkup_id
## 1 7dfa56dd7d5956b17587
## 2 4f9e6fc637eaf7b736c2
## 3 6c2545703bd527a7144d
## 4 527d6b1eaffc69ddd882
## 5 8bd62c394a35213bdf52
## 6 3b3a628f83135acd0676
##                                                      event_name
## 1 Xfinity Center Mansfield Premier Parking: Florida Georgia Line
## 2                  Gorge Camping - dave matthews band - sept 3-7
## 3                  Dodge Theatre Adams Street Parking - benise
## 4   Gexa Energy Pavilion Vip Parking : kid rock with sheryl crow
## 5                                 Premier Parking - motley crue
## 6                                     Fast Lane Access: Journey
##                        primary_act_name secondary_act_name
## 1 XFINITY Center Mansfield Premier Parking               NULL
## 2                          Gorge Camping Dave Matthews Band
## 3                          Parking Event               NULL
## 4         Gexa Energy Pavilion VIP Parking               NULL
## 5 White River Amphitheatre Premier Parking               NULL
## 6                       Fast Lane Access            Journey
##   major_cat_name      minor_cat_name la_event_type_cat
## 1          MISC              PARKING           PARKING
## 2          MISC              CAMPING           INVALID
```

1

```
## 3            MISC              PARKING          PARKING
## 4            MISC              PARKING          PARKING
## 5            MISC              PARKING          PARKING
## 6      MISC SPECIAL ENTRY (UPSELL)             UPSELL
##                                               event_disp_name
## 1 Xfinity Center Mansfield Premier Parking: Florida Georgia Line
## 2                 Gorge Camping - dave matthews band - sept 3-7
## 3               Dodge Theatre Adams Street Parking - benise
## 4   Gexa Energy Pavilion Vip Parking : kid rock with sheryl crow
## 5                             Premier Parking - motley crue
## 6                             Fast Lane Access: Journey
##
## 1    THIS TICKET IS VALID       FOR PARKING ONLY        GOOD THIS DAY ONLY       PREMIER PARKING P
## 2                                                       %OVERNIGHT C A M P I N G%* * * * *
## 3                       ADAMS STREET GARAGE%PARKING FOR 4/21/06 ONLY%DODGE THEATRE PARKING P
## 4    THIS TICKET IS VALID       FOR PARKING ONLY      GOOD FOR THIS DATE ONLY       VIP PARKING PAS
## 5                       THIS TICKET IS VALID%FOR PARKING ONLY%GOOD THIS DATE ONLY%PREMIER PAR
## 6         FAST LANE              JOURNEY              FAST LANE EVENT       THIS IS NOT A TIC
##   tickets_purchased_qty trans_face_val_amt delivery_type_cd
## 1                     1                 45          eTicket
## 2                     1                 75        TicketFast
## 3                     1                  5        TicketFast
## 4                     1                 20             Mail
## 5                     1                 20             Mail
## 6                     2                 10        TicketFast
##        event_date_time    event_dt presale_dt   onsale_dt
## 1 2015-09-12 23:30:00 2015-09-12       NULL 2015-05-15
## 2 2009-09-05 01:00:00 2009-09-04       NULL 2009-03-13
## 3 2006-04-22 01:30:00 2006-04-21       NULL 2006-02-25
## 4 2011-09-03 00:00:00 2011-09-02       NULL 2011-04-22
## 5 2005-07-31 01:00:00 2005-07-30 2005-03-02 2005-03-04
## 6 2012-07-22 02:00:00 2012-07-21       NULL 2012-04-11
##   sales_ord_create_dttm sales_ord_tran_dt   print_dt timezn_nm
## 1   2015-09-11 18:17:45        2015-09-11 2015-09-12       EST
## 2   2009-07-06 00:00:00        2009-07-05 2009-09-01       PST
## 3   2006-04-05 00:00:00        2006-04-05 2006-04-05       MST
## 4   2011-07-01 17:38:50        2011-07-01 2011-07-06       CST
## 5   2005-06-18 00:00:00        2005-06-18 2005-06-28       PST
## 6   2012-07-21 17:20:18        2012-07-21 2012-07-21       PST
##      venue_city   venue_state venue_postal_cd_sgmt_1
## 1    MANSFIELD MASSACHUSETTS                  02048
## 2       QUINCY    WASHINGTON                  98848
## 3      PHOENIX       ARIZONA                  85003
## 4       DALLAS         TEXAS                  75210
## 5       AUBURN    WASHINGTON                  98092
## 6 SAN BERNARDINO    CALIFORNIA                  92407
##            sales_platform_cd print_flg la_valid_tkt_event_flg   fin_mkt_nm
## 1 www.concerts.livenation.com         T                      N       Boston
## 2                        NULL         T                      N      Seattle
## 3                        NULL         T                      N      Arizona
## 4                        NULL         T                      N       Dallas
## 5                        NULL         T                      N      Seattle
## 6         www.livenation.com         T                      N  Los Angeles
##   web_session_cookie_val gndr_cd age_yr income_amt edu_val
```

```
## 1   7dfa56dd7d5956b17587      <NA>   <NA>        <NA>     <NA>
## 2   4f9e6fc637eaf7b736c2      <NA>   <NA>        <NA>     <NA>
## 3   6c2545703bd527a7144d      <NA>   <NA>        <NA>     <NA>
## 4   527d6b1eaffc69ddd882      <NA>   <NA>        <NA>     <NA>
## 5   8bd62c394a35213bdf52      <NA>   <NA>        <NA>     <NA>
## 6   3b3a628f83135acd0676      <NA>   <NA>        <NA>     <NA>
##   edu_1st_indv_val edu_2nd_indv_val adults_in_hh_num married_ind
## 1             <NA>             <NA>             <NA>        <NA>
## 2             <NA>             <NA>             <NA>        <NA>
## 3             <NA>             <NA>             <NA>        <NA>
## 4             <NA>             <NA>             <NA>        <NA>
## 5             <NA>             <NA>             <NA>        <NA>
## 6             <NA>             <NA>             <NA>        <NA>
##   child_present_ind home_owner_ind occpn_val occpn_1st_val occpn_2nd_val
## 1              <NA>           <NA>      <NA>          <NA>          <NA>
## 2              <NA>           <NA>      <NA>          <NA>          <NA>
## 3              <NA>           <NA>      <NA>          <NA>          <NA>
## 4              <NA>           <NA>      <NA>          <NA>          <NA>
## 5              <NA>           <NA>      <NA>          <NA>          <NA>
## 6              <NA>           <NA>      <NA>          <NA>          <NA>
##   dist_to_ven
## 1        <NA>
## 2          59
## 3        <NA>
## 4        <NA>
## 5        <NA>
## 6        <NA>
```

```r
# View column names of sales
names(sales)
```

```
##  [1] "X"                    "event_id"
##  [3] "primary_act_id"       "secondary_act_id"
##  [5] "purch_party_lkup_id"  "event_name"
##  [7] "primary_act_name"     "secondary_act_name"
##  [9] "major_cat_name"       "minor_cat_name"
## [11] "la_event_type_cat"    "event_disp_name"
## [13] "ticket_text"          "tickets_purchased_qty"
## [15] "trans_face_val_amt"   "delivery_type_cd"
## [17] "event_date_time"      "event_dt"
## [19] "presale_dt"           "onsale_dt"
## [21] "sales_ord_create_dttm" "sales_ord_tran_dt"
## [23] "print_dt"             "timezn_nm"
## [25] "venue_city"           "venue_state"
## [27] "venue_postal_cd_sgmt_1" "sales_platform_cd"
## [29] "print_flg"            "la_valid_tkt_event_flg"
## [31] "fin_mkt_nm"           "web_session_cookie_val"
## [33] "gndr_cd"              "age_yr"
## [35] "income_amt"           "edu_val"
## [37] "edu_1st_indv_val"     "edu_2nd_indv_val"
## [39] "adults_in_hh_num"     "married_ind"
## [41] "child_present_ind"    "home_owner_ind"
## [43] "occpn_val"            "occpn_1st_val"
## [45] "occpn_2nd_val"        "dist_to_ven"
```

Summarizing the data

```r
# Look at structure of sales
str(sales)
```

```
## 'data.frame':    2628 obs. of  46 variables:
##  $ X                   : chr  "1" "2" "3" "4" ...
##  $ event_id            : chr  "abcaf1adb99a935fc661" "6c56d7f08c95f2aa453c" "c7ab4524a121f9d687d2"
##  $ primary_act_id      : chr  "43f0436b905bfa7c2eec" "1a3e9aecd0617706a794" "4b677c3f5bec71eec8d1"
##  $ secondary_act_id    : chr  "b85143bf51323b72e53c" "f53529c5679ea6ca5a48" "b85143bf51323b72e53c"
##  $ purch_party_lkup_id : chr  "7dfa56dd7d5956b17587" "4f9e6fc637eaf7b736c2" "6c2545703bd527a7144d"
##  $ event_name          : chr  "Xfinity Center Mansfield Premier Parking: Florida Georgia Line" "Gor
##  $ primary_act_name    : chr  "XFINITY Center Mansfield Premier Parking" "Gorge Camping" "Parking
##  $ secondary_act_name  : chr  "NULL" "Dave Matthews Band" "NULL" "NULL" ...
##  $ major_cat_name      : chr  "MISC" "MISC" "MISC" "MISC" ...
##  $ minor_cat_name      : chr  "PARKING" "CAMPING" "PARKING" "PARKING" ...
##  $ la_event_type_cat   : chr  "PARKING" "INVALID" "PARKING" "PARKING" ...
##  $ event_disp_name     : chr  "Xfinity Center Mansfield Premier Parking: Florida Georgia Line" "Gor
##  $ ticket_text         : chr  "    THIS TICKET IS VALID        FOR PARKING ONLY        GOOD THIS DA
##  $ tickets_purchased_qty : chr  "1" "1" "1" "1" ...
##  $ trans_face_val_amt  : chr  "45" "75" "5" "20" ...
##  $ delivery_type_cd    : chr  "eTicket" "TicketFast" "TicketFast" "Mail" ...
##  $ event_date_time     : chr  "2015-09-12 23:30:00" "2009-09-05 01:00:00" "2006-04-22 01:30:00" "2
##  $ event_dt            : chr  "2015-09-12" "2009-09-04" "2006-04-21" "2011-09-02" ...
##  $ presale_dt          : chr  "NULL" "NULL" "NULL" "NULL" ...
##  $ onsale_dt           : chr  "2015-05-15" "2009-03-13" "2006-02-25" "2011-04-22" ...
##  $ sales_ord_create_dttm : chr  "2015-09-11 18:17:45" "2009-07-06 00:00:00" "2006-04-05 00:00:00" "2
##  $ sales_ord_tran_dt   : chr  "2015-09-11" "2009-07-05" "2006-04-05" "2011-07-01" ...
##  $ print_dt            : chr  "2015-09-12" "2009-09-01" "2006-04-05" "2011-07-06" ...
##  $ timezn_nm           : chr  "EST" "PST" "MST" "CST" ...
##  $ venue_city          : chr  "MANSFIELD" "QUINCY" "PHOENIX" "DALLAS" ...
##  $ venue_state         : chr  "MASSACHUSETTS" "WASHINGTON" "ARIZONA" "TEXAS" ...
##  $ venue_postal_cd_sgmt_1: chr  "02048" "98848" "85003" "75210" ...
##  $ sales_platform_cd   : chr  "www.concerts.livenation.com" "NULL" "NULL" "NULL" ...
##  $ print_flg           : chr  "T " "T " "T " "T " ...
##  $ la_valid_tkt_event_flg: chr  "N " "N " "N " "N " ...
##  $ fin_mkt_nm          : chr  "Boston" "Seattle" "Arizona" "Dallas" ...
##  $ web_session_cookie_val: chr  "7dfa56dd7d5956b17587" "4f9e6fc637eaf7b736c2" "6c2545703bd527a7144d"
##  $ gndr_cd             : chr  NA NA NA NA ...
##  $ age_yr              : chr  NA NA NA NA ...
##  $ income_amt          : chr  NA NA NA NA ...
##  $ edu_val             : chr  NA NA NA NA ...
##  $ edu_1st_indv_val    : chr  NA NA NA NA ...
##  $ edu_2nd_indv_val    : chr  NA NA NA NA ...
##  $ adults_in_hh_num    : chr  NA NA NA NA ...
##  $ married_ind         : chr  NA NA NA NA ...
##  $ child_present_ind   : chr  NA NA NA NA ...
##  $ home_owner_ind      : chr  NA NA NA NA ...
##  $ occpn_val           : chr  NA NA NA NA ...
##  $ occpn_1st_val       : chr  NA NA NA NA ...
##  $ occpn_2nd_val       : chr  NA NA NA NA ...
##  $ dist_to_ven         : chr  NA "59" NA NA ...
```

```r
# View a summary of sales
summary(sales)
```

```
##       X                 event_id          primary_act_id
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  secondary_act_id   purch_party_lkup_id  event_name
##  Length:2628        Length:2628         Length:2628
##  Class :character   Class :character    Class :character
##  Mode  :character   Mode  :character    Mode  :character
##  primary_act_name   secondary_act_name major_cat_name
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  minor_cat_name     la_event_type_cat  event_disp_name
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  ticket_text        tickets_purchased_qty trans_face_val_amt
##  Length:2628        Length:2628           Length:2628
##  Class :character   Class :character      Class :character
##  Mode  :character   Mode  :character      Mode  :character
##  delivery_type_cd   event_date_time     event_dt
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##   presale_dt         onsale_dt         sales_ord_create_dttm
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  sales_ord_tran_dt    print_dt          timezn_nm
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##   venue_city         venue_state        venue_postal_cd_sgmt_1
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  sales_platform_cd    print_flg         la_valid_tkt_event_flg
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##   fin_mkt_nm         web_session_cookie_val  gndr_cd
##  Length:2628        Length:2628             Length:2628
##  Class :character   Class :character        Class :character
##  Mode  :character   Mode  :character        Mode  :character
##     age_yr           income_amt          edu_val
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  edu_1st_indv_val   edu_2nd_indv_val   adults_in_hh_num
##  Length:2628        Length:2628        Length:2628
```

```
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  married_ind        child_present_ind  home_owner_ind
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##   occpn_val          occpn_1st_val      occpn_2nd_val
##  Length:2628        Length:2628        Length:2628
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##  dist_to_ven
##  Length:2628
##  Class :character
##  Mode  :character
```

```r
# Load dplyr
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
# Get a glimpse of sales
glimpse(sales)
```

```
## Observations: 2,628
## Variables: 46
## $ X                   <chr> "1", "2", "3", "4", "5", "6", "7", "8",...
## $ event_id            <chr> "abcaf1adb99a935fc661", "6c56d7f08c95f2...
## $ primary_act_id      <chr> "43f0436b905bfa7c2eec", "1a3e9aecd06177...
## $ secondary_act_id    <chr> "b85143bf51323b72e53c", "f53529c5679ea6...
## $ purch_party_lkup_id <chr> "7dfa56dd7d5956b17587", "4f9e6fc637eaf7...
## $ event_name          <chr> "Xfinity Center Mansfield Premier Parki...
## $ primary_act_name    <chr> "XFINITY Center Mansfield Premier Parki...
## $ secondary_act_name  <chr> "NULL", "Dave Matthews Band", "NULL", "...
## $ major_cat_name      <chr> "MISC", "MISC", "MISC", "MISC", "MISC",...
## $ minor_cat_name      <chr> "PARKING", "CAMPING", "PARKING", "PARKI...
## $ la_event_type_cat   <chr> "PARKING", "INVALID", "PARKING", "PARKI...
## $ event_disp_name     <chr> "Xfinity Center Mansfield Premier Parki...
## $ ticket_text         <chr> "   THIS TICKET IS VALID        FOR PAR...
## $ tickets_purchased_qty <chr> "1", "1", "1", "1", "1", "2", "1", "1",...
## $ trans_face_val_amt  <chr> "45", "75", "5", "20", "20", "10", "30"...
## $ delivery_type_cd    <chr> "eTicket", "TicketFast", "TicketFast", ...
## $ event_date_time     <chr> "2015-09-12 23:30:00", "2009-09-05 01:0...
## $ event_dt            <chr> "2015-09-12", "2009-09-04", "2006-04-21...
```

```
## $ presale_dt              <chr> "NULL", "NULL", "NULL", "NULL", "2005-0...
## $ onsale_dt               <chr> "2015-05-15", "2009-03-13", "2006-02-25...
## $ sales_ord_create_dttm   <chr> "2015-09-11 18:17:45", "2009-07-06 00:0...
## $ sales_ord_tran_dt       <chr> "2015-09-11", "2009-07-05", "2006-04-05...
## $ print_dt                <chr> "2015-09-12", "2009-09-01", "2006-04-05...
## $ timezn_nm               <chr> "EST", "PST", "MST", "CST", "PST", "PST...
## $ venue_city              <chr> "MANSFIELD", "QUINCY", "PHOENIX", "DALL...
## $ venue_state             <chr> "MASSACHUSETTS", "WASHINGTON", "ARIZONA...
## $ venue_postal_cd_sgmt_1  <chr> "02048", "98848", "85003", "75210", "98...
## $ sales_platform_cd       <chr> "www.concerts.livenation.com", "NULL", ...
## $ print_flg               <chr> "T ", "T ", "T ", "T ", "T ", "T ", "T ...
## $ la_valid_tkt_event_flg  <chr> "N ", "N ", "N ", "N ", "N ", "N ", "N ...
## $ fin_mkt_nm              <chr> "Boston", "Seattle", "Arizona", "Dallas...
## $ web_session_cookie_val  <chr> "7dfa56dd7d5956b17587", "4f9e6fc637eaf7...
## $ gndr_cd                 <chr> NA, NA, NA, NA, NA, NA, "M", NA, NA, NA...
## $ age_yr                  <chr> NA, NA, NA, NA, NA, NA, "28", NA, NA, N...
## $ income_amt              <chr> NA, NA, NA, NA, NA, NA, "112500", NA, N...
## $ edu_val                 <chr> NA, NA, NA, NA, NA, NA, "High School", ...
## $ edu_1st_indv_val        <chr> NA, NA, NA, NA, NA, NA, "High School", ...
## $ edu_2nd_indv_val        <chr> NA, NA, NA, NA, NA, NA, "NULL", NA, NA,...
## $ adults_in_hh_num        <chr> NA, NA, NA, NA, NA, NA, "4", NA, NA, NA...
## $ married_ind             <chr> NA, NA, NA, NA, NA, NA, "0", NA, NA, NA...
## $ child_present_ind       <chr> NA, NA, NA, NA, NA, NA, "1", NA, NA, NA...
## $ home_owner_ind          <chr> NA, NA, NA, NA, NA, NA, "0", NA, NA, NA...
## $ occpn_val               <chr> NA, NA, NA, NA, NA, NA, "NULL", NA, NA,...
## $ occpn_1st_val           <chr> NA, NA, NA, NA, NA, NA, "Craftsman Blue...
## $ occpn_2nd_val           <chr> NA, NA, NA, NA, NA, NA, "NULL", NA, NA,...
## $ dist_to_ven             <chr> NA, "59", NA, NA, NA, NA, NA, NA, NA, N...
```

Removing redundant info

```r
# Remove the first column of sales: sales2
sales2 <- sales[, -1]
```

Information not worth keeping

```r
# Define a vector of column indices: keep
keep <-c(5:30)

# Subset sales2 using keep: sales3
sales3 <- subset(sales2,select=keep)
```

Separating columns

```r
# Load tidyr
library(tidyr)

# Split event_date_time: sales4
sales4 <- separate(sales3, event_date_time,
                   c("event_dt","event_time"), sep = " ")
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 33 rows [97,
## 194, 223, 302, 466, 587, 731, 836, 945, 946, 947, 948, 1057, 1183, 1222,
## 1302, 1322, 1327, 1397, 1468, ...].
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 55 rows [87,
## 88, 195, 224, 300, 301, 303, 467, 491, 492, 530, 531, 606, 607, 649, 650,
## 695, 696, 739, 740, ...].
```

```r
# Split sales_ord_create_dttm: sales5
sales5 <-separate(sales4,sales_ord_create_dttm,c("ord_create_dt","ord_create_time"),sep = " ")
```

```
## Warning: Expected 2 pieces. Additional pieces discarded in 24 rows [97,
## 223, 491, 587, 731, 824, 946, 947, 948, 1057, 1183, 1222, 1232, 1322, 1687,
## 1860, 1985, 2024, 2189, 2239, ...].
```

```
## Warning: Expected 2 pieces. Missing pieces filled with `NA` in 64 rows [87,
## 88, 194, 195, 224, 300, 301, 302, 303, 466, 467, 492, 530, 531, 588, 606,
## 607, 649, 650, 695, ...].
```

Dealing with warnings

```r
# Define an issues vector
issues <- c(2516, 3863, 4082, 4183)

# Print values of sales_ord_create_dttm at these indices
sales3$sales_ord_create_dttm[issues]
```

```
## [1] "2016-01-19 18:10:55" NA                   NA
## [4] NA
```

```r
# Print a well-behaved value of sales_ord_create_dttm
sales3$sales_ord_create_dttm[2517]
```

```
## [1] "2006-10-14 00:00:00"
```

Identifying dates

```r
# Load stringr
library(stringr)

# Find columns of sales5 containing "dt": date_cols
date_cols <- str_detect(names(sales5),"dt")

# Load lubridate
library(lubridate)
```

```
##
## Attaching package: 'lubridate'
```

```
## The following object is masked from 'package:base':
##
##     date
```

```r
# Coerce date columns into Date objects
sales5[, date_cols] <- lapply(sales5[,date_cols], ymd)
```

```
## Warning: 32 failed to parse.
```

```
## Warning: 1550 failed to parse.
```

```
## Warning: 96 failed to parse.
```

```
## Warning: 21 failed to parse.
```

```
## Warning: 47 failed to parse.
```

```
## Warning: 243 failed to parse.
```

More warnings!

```r
# Find date columns (don't change)
date_cols <- str_detect(names(sales5), "dt")

# Create logical vectors indicating missing values (don't change)
missing <- lapply(sales5[, date_cols], is.na)

# Create a numerical vector that counts missing values: num_missing
num_missing <- sapply(missing,sum)

# Print num_missing
num_missing
```

```
##          event_dt       presale_dt        onsale_dt     ord_create_dt
##               102             1598              144                91
## sales_ord_tran_dt         print_dt
##                95              291
```

Combining columns

```r
# Combine the venue_city and venue_state columns
sales6 <- unite(sales5,venue_city_state,venue_city,venue_state,sep = ', ')

# View the head of sales6
head(sales6)
```

```
##                                                      event_name
## 1 Xfinity Center Mansfield Premier Parking: Florida Georgia Line
## 2                    Gorge Camping - dave matthews band - sept 3-7
## 3               Dodge Theatre Adams Street Parking - benise
## 4   Gexa Energy Pavilion Vip Parking : kid rock with sheryl crow
## 5                            Premier Parking - motley crue
## 6                            Fast Lane Access: Journey
##                  primary_act_name secondary_act_name
```

```
## 1 XFINITY Center Mansfield Premier Parking                NULL
## 2                       Gorge Camping Dave Matthews Band
## 3                       Parking Event                NULL
## 4        Gexa Energy Pavilion VIP Parking             NULL
## 5 White River Amphitheatre Premier Parking            NULL
## 6                       Fast Lane Access          Journey
##   major_cat_name        minor_cat_name la_event_type_cat
## 1          MISC               PARKING           PARKING
## 2          MISC               CAMPING           INVALID
## 3          MISC               PARKING           PARKING
## 4          MISC               PARKING           PARKING
## 5          MISC               PARKING           PARKING
## 6          MISC SPECIAL ENTRY (UPSELL)            UPSELL
##                                                 event_disp_name
## 1 Xfinity Center Mansfield Premier Parking: Florida Georgia Line
## 2                 Gorge Camping – dave matthews band – sept 3-7
## 3                 Dodge Theatre Adams Street Parking – benise
## 4   Gexa Energy Pavilion Vip Parking : kid rock with sheryl crow
## 5                       Premier Parking – motley crue
## 6                       Fast Lane Access: Journey
##
## 1     THIS TICKET IS VALID       FOR PARKING ONLY        GOOD THIS DAY ONLY       PREMIER PARKING P
## 2                                                   %OVERNIGHT C A M P I N G%* * * * *
## 3                             ADAMS STREET GARAGE%PARKING FOR 4/21/06 ONLY%DODGE THEATRE PARKING P
## 4     THIS TICKET IS VALID       FOR PARKING ONLY     GOOD FOR THIS DATE ONLY      VIP PARKING PAS
## 5                             THIS TICKET IS VALID%FOR PARKING ONLY%GOOD THIS DATE ONLY%PREMIER PAR
## 6        FAST LANE                 JOURNEY             FAST LANE EVENT      THIS IS NOT A TIC
##   tickets_purchased_qty trans_face_val_amt delivery_type_cd   event_dt
## 1                     1                 45          eTicket 2015-09-12
## 2                     1                 75        TicketFast 2009-09-05
## 3                     1                  5        TicketFast 2006-04-22
## 4                     1                 20             Mail 2011-09-03
## 5                     1                 20             Mail 2005-07-31
## 6                     2                 10        TicketFast 2012-07-22
##   event_time presale_dt  onsale_dt ord_create_dt ord_create_time
## 1   23:30:00       <NA> 2015-05-15    2015-09-11        18:17:45
## 2   01:00:00       <NA> 2009-03-13    2009-07-06        00:00:00
## 3   01:30:00       <NA> 2006-02-25    2006-04-05        00:00:00
## 4   00:00:00       <NA> 2011-04-22    2011-07-01        17:38:50
## 5   01:00:00 2005-03-02 2005-03-04    2005-06-18        00:00:00
## 6   02:00:00       <NA> 2012-04-11    2012-07-21        17:20:18
##   sales_ord_tran_dt   print_dt timezn_nm       venue_city_state
## 1        2015-09-11 2015-09-12       EST   MANSFIELD, MASSACHUSETTS
## 2        2009-07-05 2009-09-01       PST         QUINCY, WASHINGTON
## 3        2006-04-05 2006-04-05       MST          PHOENIX, ARIZONA
## 4        2011-07-01 2011-07-06       CST            DALLAS, TEXAS
## 5        2005-06-18 2005-06-28       PST         AUBURN, WASHINGTON
## 6        2012-07-21 2012-07-21       PST SAN BERNARDINO, CALIFORNIA
##   venue_postal_cd_sgmt_1         sales_platform_cd print_flg
## 1                  02048 www.concerts.livenation.com        T
## 2                  98848                       NULL        T
## 3                  85003                       NULL        T
## 4                  75210                       NULL        T
## 5                  98092                       NULL        T
```

```
## 6                   92407        www.livenation.com        T
##   la_valid_tkt_event_flg  fin_mkt_nm
## 1                      N        Boston
## 2                      N       Seattle
## 3                      N       Arizona
## 4                      N        Dallas
## 5                      N       Seattle
## 6                      N  Los Angeles
```