

# Proгноза потребления газа природного

# Problema

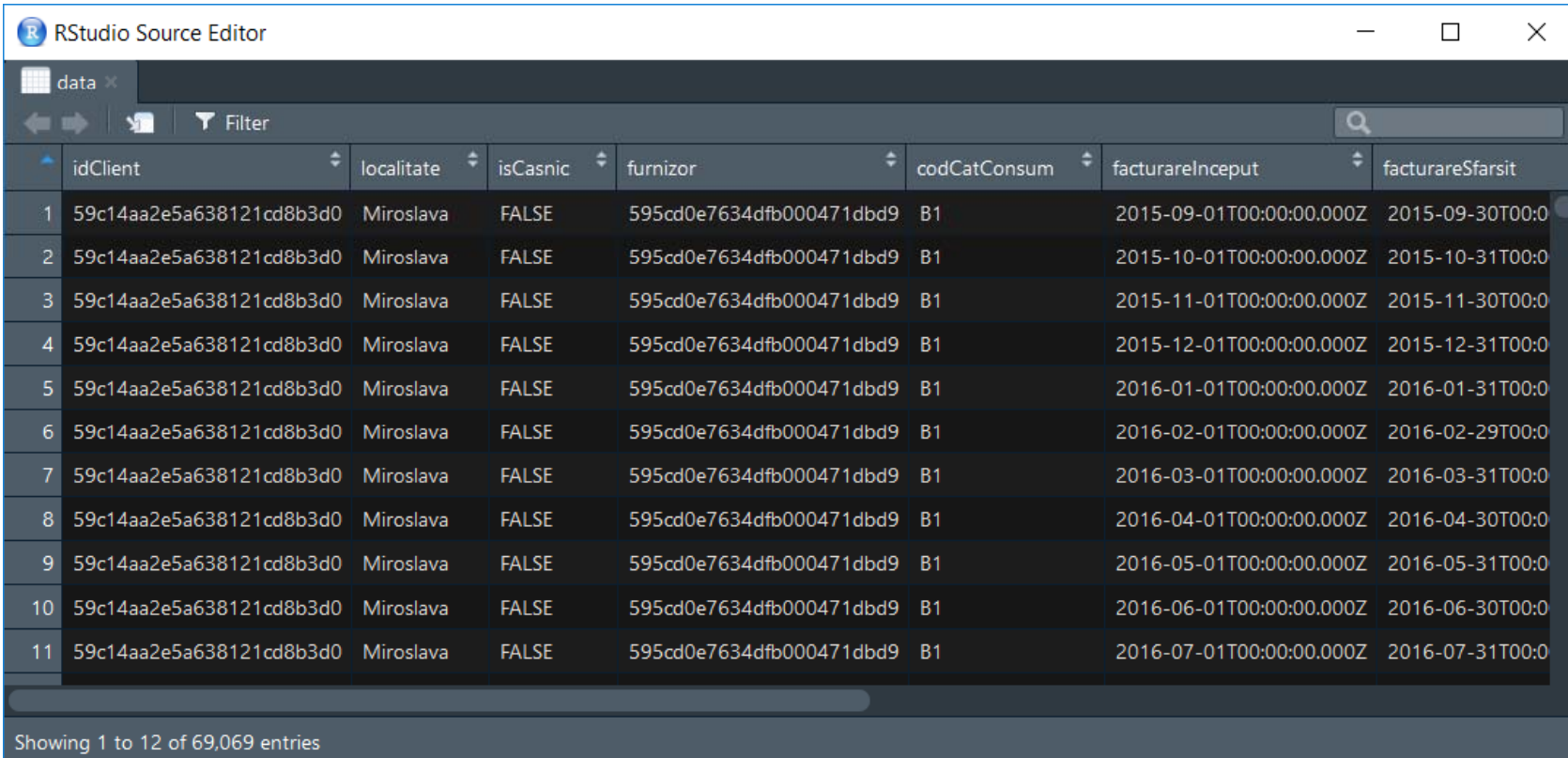
- Consum de gaz
- Furnizor
- Distribuitor – rezervă gaz la furnizor pt. clienți
- Client
- Anual, trimestrial, lunar, zilnic



CÂT?

# Datele inițiale (1)

- 14 coloane
- 69 069 rânduri
- Sept. 2015 – Nov. 2017
- Cât a consumat fiecare client într-o lună

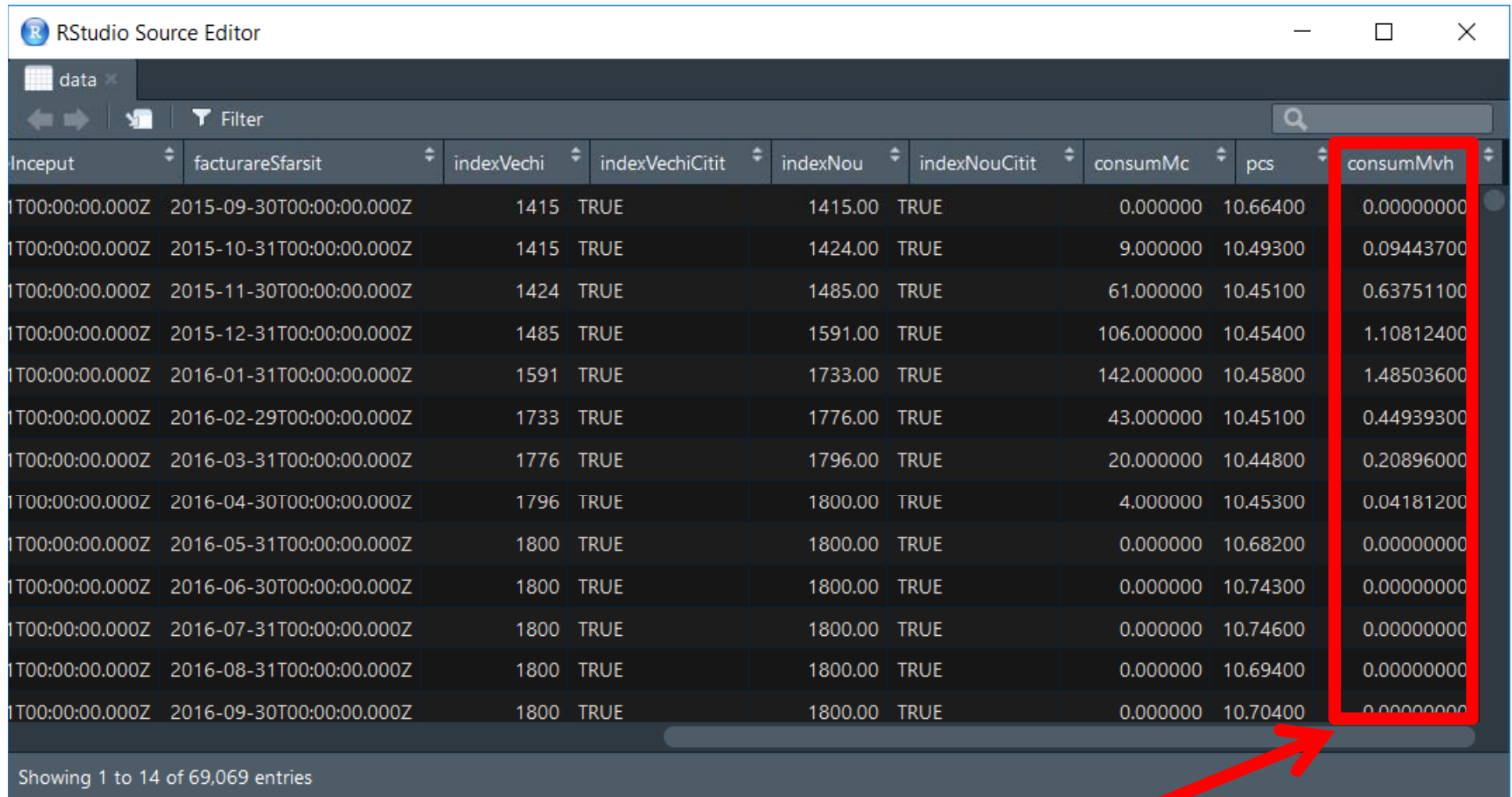


The screenshot shows the RStudio Source Editor window with a data table. The table has 8 columns: idClient, localitate, isCasnic, furnizor, codCatConsum, facturareInceput, and facturareSfarsit. The first 11 rows are visible, showing data for a client in Miroslava. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2015-09-01 and ends on 2015-09-30. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2015-10-01 and ends on 2015-10-31. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2015-11-01 and ends on 2015-11-30. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2015-12-01 and ends on 2015-12-31. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2016-01-01 and ends on 2016-01-31. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2016-02-01 and ends on 2016-02-29. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2016-03-01 and ends on 2016-03-31. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2016-04-01 and ends on 2016-04-30. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2016-05-01 and ends on 2016-05-31. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2016-06-01 and ends on 2016-06-30. The status is FALSE, and the provider is 595cd0e7634dfb000471dbd9. The consumption code is B1. The billing period starts on 2016-07-01 and ends on 2016-07-31.

	idClient	localitate	isCasnic	furnizor	codCatConsum	facturareInceput	facturareSfarsit
1	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2015-09-01T00:00:00.000Z	2015-09-30T00:00:00.000Z
2	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2015-10-01T00:00:00.000Z	2015-10-31T00:00:00.000Z
3	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2015-11-01T00:00:00.000Z	2015-11-30T00:00:00.000Z
4	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2015-12-01T00:00:00.000Z	2015-12-31T00:00:00.000Z
5	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2016-01-01T00:00:00.000Z	2016-01-31T00:00:00.000Z
6	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2016-02-01T00:00:00.000Z	2016-02-29T00:00:00.000Z
7	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2016-03-01T00:00:00.000Z	2016-03-31T00:00:00.000Z
8	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2016-04-01T00:00:00.000Z	2016-04-30T00:00:00.000Z
9	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2016-05-01T00:00:00.000Z	2016-05-31T00:00:00.000Z
10	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2016-06-01T00:00:00.000Z	2016-06-30T00:00:00.000Z
11	59c14aa2e5a638121cd8b3d0	Miroslava	FALSE	595cd0e7634dfb000471dbd9	B1	2016-07-01T00:00:00.000Z	2016-07-31T00:00:00.000Z

Showing 1 to 12 of 69,069 entries

# Datele inițiale (2)



RStudio Source Editor

data

Filter

Inceput	facturareSfarsit	indexVechi	indexVechiCitit	indexNou	indexNouCitit	consumMc	pcs	consumMvh
2015-09-30T00:00:00.000Z	2015-09-30T00:00:00.000Z	1415	TRUE	1415.00	TRUE	0.000000	10.66400	0.00000000
2015-10-31T00:00:00.000Z	2015-10-31T00:00:00.000Z	1415	TRUE	1424.00	TRUE	9.000000	10.49300	0.09443700
2015-11-30T00:00:00.000Z	2015-11-30T00:00:00.000Z	1424	TRUE	1485.00	TRUE	61.000000	10.45100	0.63751100
2015-12-31T00:00:00.000Z	2015-12-31T00:00:00.000Z	1485	TRUE	1591.00	TRUE	106.000000	10.45400	1.10812400
2016-01-31T00:00:00.000Z	2016-01-31T00:00:00.000Z	1591	TRUE	1733.00	TRUE	142.000000	10.45800	1.48503600
2016-02-29T00:00:00.000Z	2016-02-29T00:00:00.000Z	1733	TRUE	1776.00	TRUE	43.000000	10.45100	0.44939300
2016-03-31T00:00:00.000Z	2016-03-31T00:00:00.000Z	1776	TRUE	1796.00	TRUE	20.000000	10.44800	0.20896000
2016-04-30T00:00:00.000Z	2016-04-30T00:00:00.000Z	1796	TRUE	1800.00	TRUE	4.000000	10.45300	0.04181200
2016-05-31T00:00:00.000Z	2016-05-31T00:00:00.000Z	1800	TRUE	1800.00	TRUE	0.000000	10.68200	0.00000000
2016-06-30T00:00:00.000Z	2016-06-30T00:00:00.000Z	1800	TRUE	1800.00	TRUE	0.000000	10.74300	0.00000000
2016-07-31T00:00:00.000Z	2016-07-31T00:00:00.000Z	1800	TRUE	1800.00	TRUE	0.000000	10.74600	0.00000000
2016-08-31T00:00:00.000Z	2016-08-31T00:00:00.000Z	1800	TRUE	1800.00	TRUE	0.000000	10.69400	0.00000000
2016-09-30T00:00:00.000Z	2016-09-30T00:00:00.000Z	1800	TRUE	1800.00	TRUE	0.000000	10.70400	0.00000000

Showing 1 to 14 of 69,069 entries

De prezis...

# Probleme principale

- În preprocesarea datelor
- În “modul de vizualizare” a datelor  
(Cum privim datele?)

# Probleme în preprocesarea... (1)

- Pentru unii clienți avem mai multe citiri de contor în aceeași lună
- Pentru unii clienți nu avem o citire reală de contor (indexNouCitit = FALSE)
- Schimb de id de la o lună la alta
- Să normalizăm datele?

# Probleme în preprocesarea... (2)

- Din cele 14 coloane:
  - 1 este de prezis
  - Doar 3 coloane care ne-ar ajuta la prezis și **niciuna nu este numerică**:
    - data
    - codCatConsum
    - isCasnic
  - Este îndeajuns?

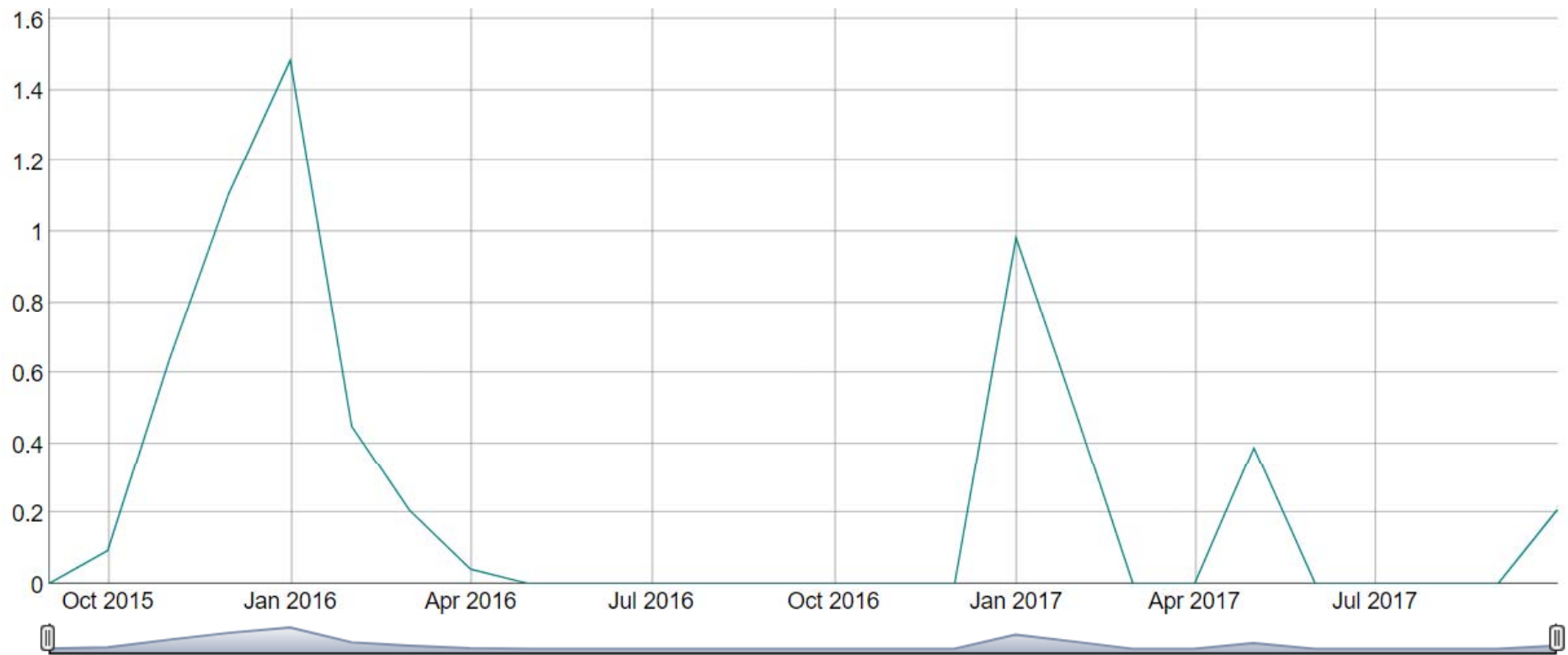
# Probleme în preprocesarea... (3)

- ++: consumuri din lunile anterioare
- ++: date despre temperatură
  - Prognoza doar pe 10 zile
  - DAR distribuitorul trebuie să rezerve până pe data de 10 a lunii anterioare...



# Probleme: Cum privim datele?

- Model global - lună: punem luna ca atribut de intrare
- Model *local* - lună: un model pentru fiecare lună
- Model global – client: toți clienții la un loc => modelul
- Model local – client: un client => modelul  
(serii de timp)



# Probleme: Organizarea datelor

- Date puține

Train set	Validation/Development set	Test set
2016-11-01 <i>[nu pentru serii de timp]</i>	2017-11-01	-
2016-10-01 <i>[nu pentru serii de timp]</i>	2017-10-01	-
2015-09-01 to 2016-11-01	2016-12-01	Posibil Decembrie 2017
2015-09-01 to 2016-11-01	2017-01-01	Posibil Ianuarie 2018
2015-09-01 to 2016-12-01	2017-01-01	Posibil Ianuarie 2018
2015-09-01 to 2016-11-01	2017-02-01	-
2015-09-01 to 2017-01-01	2017-02-01	-
2015-09-01 to 2016-11-01	2017-03-01	-
2015-09-01 to 2017-02-01	2017-03-01	-
2015-09-01 to 2016-11-01	2017-04-01	-

...

# Clienții

	Miroslava	Panciu	Chirnogi	Odoboești	Total
B1-NonCasnic	117	46	14	27	<b>204</b>
B1-Casnic	1690	911	267	631	<b>3499</b>
B2-NonCasnic	19	20	0	12	<b>51</b>
B2-Casnic	121	37	9	38	<b>205</b>
B3-NonCasnic	6	15	0	5	<b>26</b>
B3-Casnic	0	1	0	6	<b>7</b>
B4-NonCasnic	1	0	0	2	<b>3</b>
B4-Casnic	0	0	0	1	<b>1</b>
<b>Total</b>	<b>1954</b>	<b>1030</b>	<b>290</b>	<b>722</b>	<b>3996</b>

Tipuri de clienți [doar cei <b>casnici</b> au fost considerați]	
Din Miroslava – Corelați cu temperatura	Din Miroslava – Necorelați cu temperatura
Din Panciu – Corelați cu temperatura	Din Panciu – Necorelați cu temperatura
Din Chirnogi – Corelați cu temperatura	Din Chirnogi – Necorelați cu temperatura
Din Odoboești – Corelați cu temperatura	Din Odoboești – Necorelați cu temperatura

# Modele folosite (1)

- Nereferitoare la serii de timp
  - Regresia liniară  $y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon$
  - ANN
    - MLP \_-50-50-1
    - MLP \_-30-20-1
- Atribute de intrare:
  - consumuri anterioare și
  - tipul clientului [B1, B2, B3, or B4]: cum? și/sau
  - luna în care se prezice consumul: cum? și/sau
  - diferențe anuale și/sau
  - temperatura în luna în care se prezice consumul

# Modele folosite (2)

- Referitoare la serii de timp
  - Modelul naiv = de acum 1 an
  - ARIMA      AR:  $y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \varepsilon_t$
  - ARIMAX      MA:  $n_t = \alpha_0 + \alpha_1 \varepsilon_{t-1} + \dots + \alpha_q \varepsilon_{t-q} + \varepsilon_t$   
 $y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_k y_{t-p} + n_t$
- Atribute de intrare:
  - consumuri anterioare și
  - temperaturi anterioare
  - ... Și hiper-parametri diferiți:
    - lambda, D

# Metrici

$$MSE = \text{mean}((y - \hat{y})^2) = \frac{\sum (y - \hat{y})^2}{n}$$

$$MAE = \text{mean}(|y - \hat{y}|) = \frac{\sum |y - \hat{y}|}{n}$$

$$MdAE = \text{median}(|y - \hat{y}|)$$

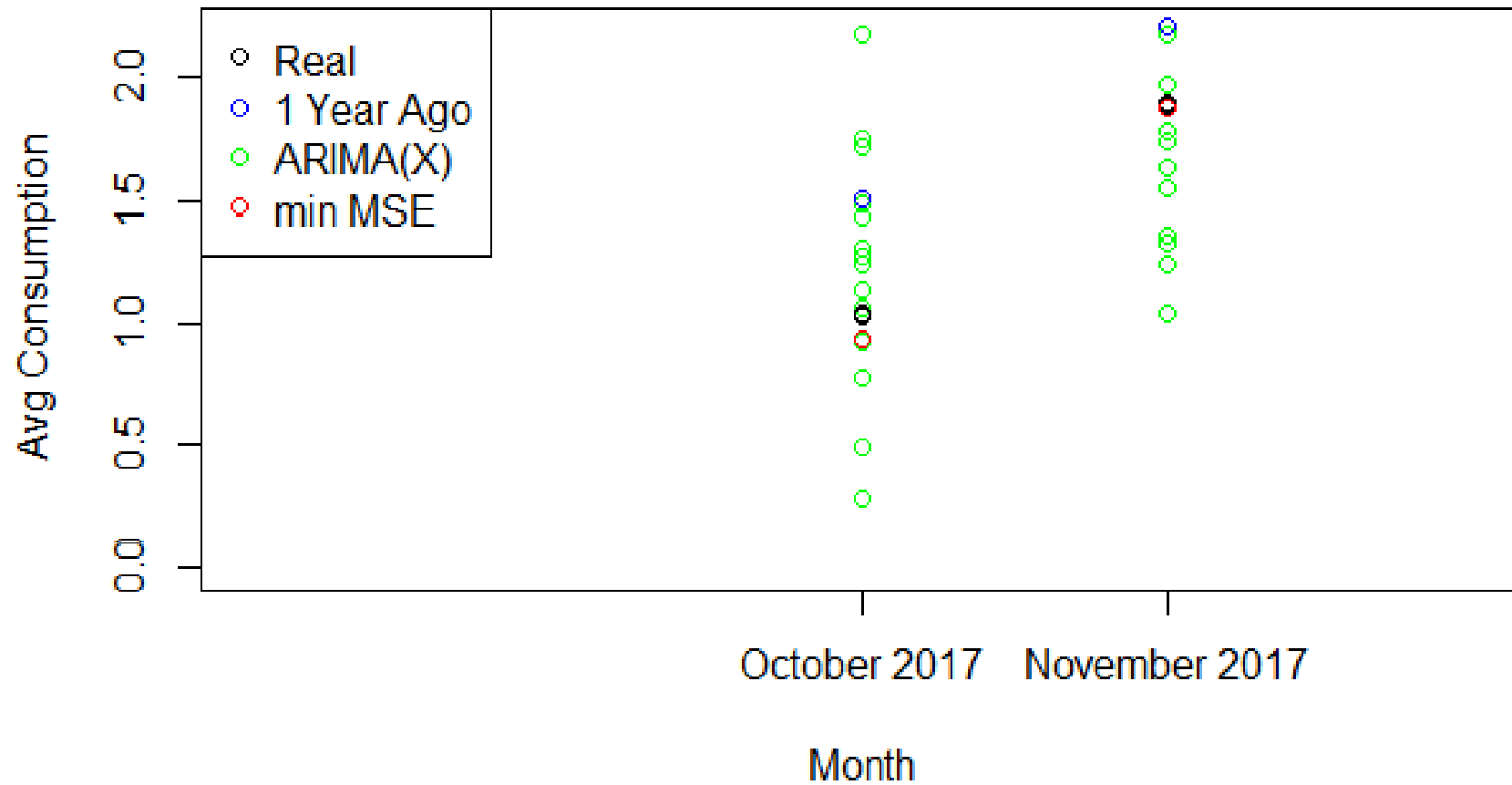
$$Corr = \text{cor}(y, \hat{y})$$

Diferența absolută între consumul mediu prezis  
și consumul mediu real

Model	MSE	MAE	MdAE	Corr	Abs. Avg. Diff.
Validare în Decembrie 2016					
<b>MLP_-50-50-1</b> <b>withTemp: N</b> Train Period: 2015-09 to 2016-11 #Lags: 13 typeMonth: one hot typeClient: one hot withDiff: Y #Runs: 1 #Records: 765	0.3798351	0.4314502	0.3268276	0.9227971	0.234341900
<b>MLP_-50-50-1</b> <b>withTemp: Y</b> Train Period: 2015-09 to 2016-11 #Lags: 13 typeMonth: one hot typeClient: none withDiff: Y #Runs: 1 #Records: 765	0.3801115	0.4244243	0.3111872	0.9167776	0.049795906
Validare în Ianuarie 2017					
<b>Linear regression</b> <b>withTemp: N</b> Train Period: 2015-09 to 2016-11 #Lags: 13 typeMonth: number typeClient: one hot withDiff: Y #Runs: 1 #Records: 796	0.6465580	0.5723214	0.4344356	0.9220211	0.1732643772
<b>Linear regression</b> <b>withTemp: Y</b> Train Period: 2015-09 to 2016-11 #Lags: 13 typeMonth: number typeClient: one hot withDiff: Y #Runs: 1 #Records: 796	0.6465580	0.5723214	0.4344356	0.9220211	0.1732643772



## Rezultate (2)



# Concluzii

- Temperatura ajută doar pentru 3 luni: mai, iulie și octombrie
- Pe vară: regresia liniară
- Modelul naiv nu este cel mai bun în nicio lună
- Modelele cu serii de timp: **X**