18f0371 talha Zafar

```
raw_img %>%
  image_crop(geometry_area(350, 0,350, 10)) %>%
  ocr()

## [1] "ye May 5, 2019\n|\n[I =\nRe/Liter\n[| 73.407\ncease 72.001\n720\n73a0 | 66.06 |\n7326 |\nMOGAS E510 Gasoline\nRetail_| Retail | Direct\n73.26 | 73.26\n[329 | ose| 056\n(76.69 | 73.80 | 73.80\n| SeR=oee | EOE |=\n|_76.69 | 73.80 | 73.80\n_ Ser ber ee| 264\ni007] a6 | 809\nite ae |\n[ez | 15.39) 1539\n108.42 Lasse 1085 |\n@sn\nper by age. [231%\n"
```

Including Plots

```
raw_img %>%
  image_quantize(colorspace = "gray") %>%
  image_ggplot()
```

E-10 Gasoline Price effective May S	2019		
		Fa/Liter	
A HOCAS I SO Michael Alvenge cost of supply		79.43	
5 Diseased Pro-Distributy prints in a work of by 1900		77.00	
C 18(4) part (Glass of Kir Dirit Europe to (1996) of (6)		2.20	
D 54 % part Mogas Es-Fit Energy price (20% of a)	12.41	in to	
C to se seem embegs + mg		73.2%	
Max. Ex-Tapo: Sale Price	MEGAF	K10C	enline
	Tetal)	1ms	Divin
FBI Sed of Excepty / H-12 ExcePt Le	72.40	72.00	110
UIHP.	2.72	3.54	1.9
H[Substants](F+G)	76.69	73.00	73.00
Price Hilleren Hall Chrise	2023	0.0	10.00
Substotal (III + I)	20.07	72.83	77.80
N OKCO Hargin	2.64	2.94	204
Dealers Commission	1.47	5.47	
M February Laws as northed by Federal Sova.	74 (1)	2.142	-34.05
M Submorté () – K + L + M)	96.04	98.50	98.53
Olemetre	1142	15.00	2,524
Max Ex-Lepo; Sale (vice (N + 0)	18842	305.94	105.53
Difference between MCGaS and CID-Ro/Alfrer)	(250)	W. 8005	
E 16 Chapterly Dags		2.31%	

```
fuzz_fun <- function(fuzz){
  raw_img %>%
   image_quantize(colorspace = "gray") %>%
   image_transparent(color = "white", fuzz=fuzz) %>%
   image_background("white") %>%
   image_crop(geometry_area(350, 0,350, 10))
}
```

fuzz_fun(20)

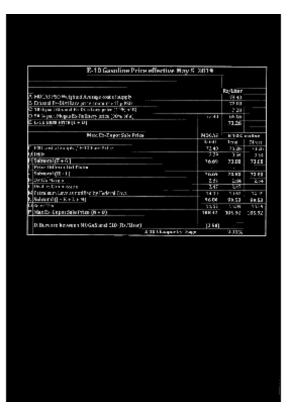
ve May 5,	2019		
		Rs/Liter	
		73.40	
		72.00	
		7.20	
	73.40	66.06	
		73.26	
	MOGAS	E-10 Ga	soline
	Retail	Retail	Direct
	73.40	73.26	73.26
	3.29	0.54	0.54
	76.69	73.80	73.80
	76.69	73.80	73.80
	2.64	2.64	2.64
	3.47	3.47	
	14.00	10.62	14.09
	96.80	90.53	90.53
	11.62	15.39	15.39
	108.42	105.92	105.92
	(2.50)	-	
per by %age	, ,	2.31%	

```
no_grid <- raw_img %>%
  image_quantize(colorspace = "gray") %>%
  image_transparent(color = "white", fuzz=20) %>%
  image_background("white")

image_ggplot(no_grid)
```

E-10 Gasoline Price effective P	lay 5, 2019	Г	
		Ro/Liber	
A MOCASTSO Multiplicad Average coat of supply		79.40	
5. Drivated Re-Distribute price many medity (80)		72.00	
C Tillebyon (Blocked Broth of Large police (C 199) of E)		7.20	
O 54 (e-pair, Mugas Es-Fellinery, price (10% of a)	12.44	66.56	
E to the mean entrape + or		73.26	
Max. Co-Zapor: Sale Price	MacAs	ктас	roller
•	Bainti	Tenni	Dixm
PSS and allowards (H13 Farm Fell y	72.40	73.25	110
SIRP	2.72	3.94	7.9
· [Submorn I (F + G)	16.69	73.00	73.00
Price Hitleren Hat Chairs			-
Substore 1 (21 + 1.)	*0.00	72.83	22.60
OMCS Margin	2.51	2,64	2.04
Biodins Commission	3.47	5/42	
e Fished man Larve as notified by Federal Cova.	14.31	1.162	72.7
Subscrib() - K + 1 + M)	96.00	98.53	98.57
Milde on The	33.52	1100	1324
Mac Es - Depor Solo Frico (N + 0)	18842	305.94	005.55
D.Barung between NOGaS and CID (Rt./Hinr)	(2.64)	. —	
£31 thagart		2.31%	

```
no_grid %>%
  image_negate() %>%
  image_ggplot()
```



```
no_grid %>%
  image_negate() %>% # negate
  image_morphology(method = "Thinning", kernel = "Rectangle:20x1") %>%
  image_negate() %>% # back to white
  image_ggplot()
```

E-10 Gasoline Price effective	May 5 2019	Γ	
		RayLitar	
A MOCASINO Multiplicad Average cost of supply		79.40	
5. Diagral Re-Distillary price many medity (80)		72.00	
Contribution (Richard Re Distribute) police (CON) e10)		7.20	
O 54 Separt Mogas Ex-Relinery price (90% of a)	12.44	66.56	
Erreram sutatr+ of		71.26	
Mac. Co-Capor Sale Price	MOGAN	ктас	cultur
	Brieff	1773	Dixn
PSC and allow pix (1913 Familia) in	72.40	73.00	110
3 He>	7.75	3.54	7.5
F[Subrocol(F+G)]	76.69	73.00	73.0
Principalities and the Chairs			-
Substoral (III + I)	70.07	72.83	77.80
Series Hang a	2.51	2,64	204
Block to Contractors	1.47	3/47	
M February Larve as notified by Federal Styr.	14.31	1.162	12.75
5 Salastri() - K+1 + M;	96.00	98.50	94.52
Modern The	33.52	1,528	7324
Mac Es-Copo: Sale Price (N + 0)	18842	305.94	305.57
D.Burung between NOGaS and CID (R)/Hine)	(250)	. —	
D31 thags		2.31%	

no_grid %>%
 image_crop(geometry_area(350, 0,350, 10)) %>%
 image_ggplot()

ve May 5, 2019		
	Re/litter	
	75.40	
	72.00	
	7,30	
70.40	\$500	
	75.24	
MO245	R-101c	990 EBA
fr at	90,000	Local
75.4E	71.26	7126
C-04	0.64	754
76.63	73/89	73.00
76.63		73.30
1.64	2.64	764
2.47	2.47	
14.00	10.00	14.05
9680	19175	90.53
11.62	15.39	1539
10645	10596	105.12
(6.50))	
parties the cal	3.5.1%	

```
no_grid_crop <- no_grid %>%
  image_crop(geometry_area(350, 0,350, 10))
no_grid_crop %>%
  image_ggplot()
```

		_	
ve May 5. 2	2019	_	
		No. 1115	
		Rs/titles 75.40	-
		73.00	_
		7,000	
	70,40	25000	
		79.24	
	MOSAS	16-1016	
	te at	October	Low
	75,40	71.26	7126
	4.04	0.64	754
	76.63	73/89	73.00
	76.63	73390	73.30
	1.64	2.64	764
	2.47	2.47	
	14.00	10.00	14.05
	9883	19170	90.53
	11.62	15.39	1539
	10843	10595	105.12
	(3.50)	_	
partie for out		2-51%	

```
no_grid_crop %>%
  image_ocr()
## [1] "ve May 5, 2019\n-\n[I\nRe [tier\n[7340 |\n[200\n720\nTaD 66.06\n73.26]
\nMOGAS \"~F:40 Castine\nRetail | Reta | Direc\n73.26 | 73.26\n[3297 054
054\n[7669 | 73.20 | 7380\nTt\n[7669 | 73.90 | 73.90\n- Ser der ee|\n-
349\n~~ rae |e\n360 [9053 | 9053\nsae | weet 15.39\n108.42 | 105.92 | 105.92
\nesa\nper by age 23%\n"
num_only <- tesseract::tesseract(</pre>
 options = list(tessedit_char_whitelist = c(".0123456789 "))
no_grid %>%
  image_quantize(colorspace = 'gray') %>%
  image_threshold() %>%
  image_crop(geometry_area(350, 0,350, 10)) %>%
  ocr(engine = num_only)
## [1] "5 2019\n73.40\n72.00\n7 320\n730 66.06\n73.26\n0 10\n73407326
73.26\n329 054 054\n76.69 73.80 73.80\n76697380 73.80\n264 264 264\n37
```

```
347\n14.00 1062 1408\n9680 9053 90.53\n11.62 15391530\n408.42 105.92
10592\n50\n. 2314.\n"
combo <- tesseract::tesseract(</pre>
  options = list(
    tessedit char whitelist = paste0(
      c(letters, LETTERS, " ", ".0123456789 (-)"), collapse = "")
  )
)
raw_text <- no_grid %>%
  image_quantize(colorspace = "gray") %>%
  image_transparent("white", fuzz = 22) %>%
  image background("white") %>%
  image threshold() %>%
  image_crop(geometry_area(300, 0,400, 10)) %>%
  ocr(engine = combo)
raw_tibble <- raw_text %>%
  str_split(pattern = "\n") %>%
  unlist() %>%
  tibble(data = .)
raw_tibble$data=gsub(')','',as.character(raw_tibble$data))
raw_tibble$data=gsub('-','',as.character(raw_tibble$data))
col1=raw tibble[1:7,]
raw tibble=raw tibble[-(1:9),]
raw_tibble=raw_tibble[-(11:13),]
col1=col1[-(1:2),]
col1=sub('.*? ','',col1$data)
col1=data.frame(col1)
check=separate(raw_tibble,data ,into = c('retail1','retail2','direct'),sep =
'\\ ',extra = 'merge')
## Warning: Expected 3 pieces. Missing pieces filled with `NA` in 1 rows [6].
library(tidyverse)
col1=col1 %>% add row(col1=check$retail2)
col1
##
        col1
## 1
       7340
      72.00
## 2
## 3
          20
## 4
       66.06
## 5
      73.26
## 6
       7326
## 7
        054
## 8 7320
```

```
## 9 7380

## 10 264

## 11 347

## 12 10627

## 13 9053

## 14 1539

## 15 105.92
```

Note that the echo $\,=\,$ FALSE parameter was added to the code chunk to prevent printing of the R code that generated the plot.