Galaxy Zoo: Morphological Classifications for Galaxies in HST Legacy Imaging

Lead Author and other Galaxy Zoo science team members

* This publication has been made possible by the participation of more than 200,000 volunteers in the Galaxy Zoo project. Their contributions are individually acknowledged at `http://authors.galaxyzoo.org/authors.html`.

E-mail: lead.author@university.edu

5 August 2015

ABSTRACT

This will be the data release paper for GZ Hubble. We present the classifications, the methodology for data reduction and corrections for redshift dependent biases in the observed morphologies.

1 INTRODUCTION

2 SAMPLE AND DATA

2.1 Summary of HST Legacy Survey Imaging

• AEGIS has 1 orbit each (2200 seconds) in F606W (V band) and F814W filter and has been dithered to 0.03 "/pixel; the imaging covers ~ 710 arcmin^2.

• GOODS targeted 2 fields, GOODS-N and GOODS-S, imaging in 4 filters – F435W (B), F606W (V), F775W (i), and F850LP (z). The mean exposure times vary, from 1000 - 2100 seconds. The images have been dithered to a pixel scale of 0.03 "/pixel and covers at total area of 320 arcmin^2 (160 arcmin^2 per field). The filters that Griffith et al. (2012) uses for the colored images were F606W and F775W for GOODS-N and F606W and F850LP for GOODS-S.

• COSMOS has 1 orbit (2028 seconds) in the F814W (I band) filter and has been dithered to 0.05 "/pixel; it covers the largest area, 1.8 deg^2.

• GEMS has 1 orbit (2160 and 2286 seconds) in the F606W and F850LP filters, with a pixel scale of 0.03 "/pixel; it covers 800 arcmin^2

2.2 User Weighting

The votes of individual users who classified galaxies in Galaxy Zoo Hubble are combined to make a vote fraction for each question on the classification tree. Users votes are weighted slightly (in a method identical to that described in Willett et al. 2013) such that users who frequently disagree with all other users end up having very low weights. The majority of users have weights very close to 1.0 (**STEVEN: Is this true for GZH - do you have a plot of the distribution of user weights or consistencies we can include here?**.

3 CORRECTING FOR REDSHIFT DEPENDENT CLASSIFICATION BIAS

Previous version of Galaxy Zoo morphology classifications (Lintott et al. 2008; Willett et al. 2013) were based on observations of galaxies in the Sloan Digital Sky Survey (SDSS) which are typically at $z < 0.1$. In these cases it could be assumed that there was no real evolution of the morphologies of galaxies and therefore any observed changes in the distribution of galaxies with different consensus morphologies was assumed to be due to the effects of redshift on the image quality (*i.e..* the reduction in physical resolution, surface brightness dimming etc). So for both previous releases of GZ morphologies, a correction was provided for redshift dependent bias based on matching the classification fractions at the highest redshfts with those at the lowest redshift. See Bamford et al. (2009) and Willett et al. (2013) for the details.

In the GZH samples the redshift range is so large that we expect to find redshift evolution of the types and morphologies of galaxies which are seen. So the previous methods of correcting for redshift dependent bias will not work. In addition the effects of band shifting will change the images even more across these redshift ranges. Figure **??** which illustrates one possible effect of losing features in spiral galaxies at high redshift).

In order to test the effects of redshift we made images of the same galaxy at a variety of redshifts using input images from the SDSS (York et al. 2000; Strauss et al. 2002) and the FERENGI code (Barden et al. 2008) to produce images that match the observational properties of those observed in the HST surveys.

3.1 Selection of FERENGI Input Galaxies

We selected 288 unique galaxies from SDSS imaging to run through the FERENGI code.

The selection spanned a variety of galaxy morphologies

Table 1. Summary of GZ Hubble Samples

Survey	$t_{\exp}$ seconds	Filters	Resolution $''$/pix	Area sq arcmin	N_{galaxies}
AEGIS	2200	F606W (V) and F814W (I)	0.03	710	
COSMOS	2028	F814W (I)	0.05	6480	
GEMS	2160/2286	F606W (V) and F850LP (z)	0.03	800	
GOODS	1000-2100s	F435W (B), F606W (V), F775W (i), F850LP (z)	0.03	320	
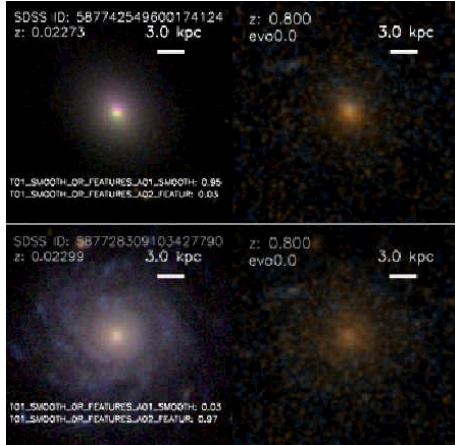
Figure 1. Examples of an obvious spiral and obvious elliptical at $z = 0$ which both look similar when redshifted to $z = 0.8$. See below for the description of the method for redshifting.

(as indicated by GZ2 classifications), with a range of surface brightnesses, and also spanned the redshift range of SDSS targets (in $N_z = 4$ bins) in order to be optimised for different target minimum redshifts in HST imaging.

The selection criteria for the different morphological categories is summarised in Table 2. The surface brightness selection ($N_\mu = 3$) was (1) low: $\mu > 21.5 \text{magarsec}^2$; (2) mid: $20.5 < \mu < 21.5 \text{magarsec}^2$; and (3) high: $\mu < 20.5 \text{magarsec}^2$.

For each of the four "target redshifts" ($z = 0.3, 0.5, 0.8$ and 1.0), the images were redshifted in $\Delta z = 0.1$ bins up to $z = 1.0$. In addition different evolution models were assumed in FERENGI. This evolution is a crude mechanism that is suppose to mimic the brightness increase of galaxies with increasing redshift. The evolution mechanism allows us to make galaxies brighter as linear function of redshift, and that is it. It is an empirical addition to the magnitude of a galaxy of the form $M' = e \times z + M$, where M' is the corrected magnitude, e is the evolutionary correction in magnitudes. (i.e. $e = -1$ essentially brightens the galaxy by 1 magnitude by $z = 1$). We ran FERENGI for values of e starting from $e = 0$ and decreasing to $e = -3.5$ in $\Delta e = 0.5$ steps. Depending on the target, or starting redshift we ran FERENGI for several final redshifts and different evolution models as summarized in Table 3.

Two series of example redshifted images are shown in Figure 2, both for runs with no evolution ($e = 0$).

Table 3. Summary of FERENGI runs

z_{target}	$N_{z\text{bins}}$	$N_{\text{evolution}}$	$e_{\max}$	N_{runs}	N_{total}
0.3	8	7	-3.0	56	4031
0.5	6	4	-1.5	24	1728
0.8	3	3	-1.0	9	648
1.0	1	3	-1.0	3	216

3.2 Approach to debiasing morphological classifications

The approach used in GZH for correcting the weighted classifications for user bias rests on the assumption that the *amount* of bias is a function of the apparent size and brightness of the image as seen on screen. This is controlled by two types of parameters: **intrinsic** properties of the galaxy itself, such as its physical diameter and luminosity, and **extrinsic** properties, such as the distance (redshift) of the galaxy and its relative orientation. There are likely other parameters that affect user accuracy, such as the proximity of close companions ("distraction bias"; see Johnson et al. 2015) or those that vary by individual user. The combination of all such parameters forms a high-dimensional space, and we have insufficient test galaxies to measure the individual effects of each. Instead, we use just two parameters that are intended to capture the bulk of the change in bias (based on GZ1/GZ2): a galaxy's r-band surface brightness (μ_r; intrinsic) and redshift (z; extrinsic).

The change in bias as a function of μ_r and z is measured using the FERENGI images at all evolutionary corrected factors. We assume that the "true" (ie, debiased) vote fraction $f_{\mu,z}$ for a galaxy can be expressed as:

$$f_{\mu,z} = (f_{\mu,z=0.3}) \times e^{\frac{z-z_0}{\zeta}}, \tag{1}$$

where $f_{\mu,z=0.3}$ is the "calibrated" vote fraction at the lowest redshift in the FERENGI bins, which is at redshift $z = 0.3$ and ζ is a positive parameter that controls the rate at which f drops off with increasing redshift. This formula fits the data relatively well (with almost no exceptions, the vote fractions for featured galaxies decrease monotonically with increasing redshift), and the exponential function bounds the observed vote fractions between $f_{\mu,z=0.3}$ and zero. Figure 3 show examples of the change in vote fraction and their fits to Equation 1 for a random selection of galaxies in the FERENGI images.

We then take the values of ζ for *all* sets of artificially redshifted galaxies and use them to fit the overall distribution of ζ as a function of surface brightness (since we expect the correction being applied to vary as a function of the intrinsic galaxy properties). We restrict the galaxies that can

Table 2. Summary of morphological categories selected for FERENGI sample. ==For Edmond: does $p_{\text{obvious/dominant}} > 0.5$ mean $(p_{\text{obv}} + p_{\text{dom}} > 0.5)$ or $(p_{\text{ob}}$ OR $p_{\text{dom}} > 0.5)$?==

Morphology	Label	Selection	N_{objects} $[N_z \times N_\mu]$
Features	Yes	$p_{\text{features}} > 0.8$, $p_{\text{odd}} < 0.1$	12
	Int	$0.3 < p_{\text{smooth}} < 0.6$, $p_{\text{odd}} < 0.1$	12
	No	$p_{\text{smooth}} > 0.8$, $p_{\text{odd}} < 0.1$	12
Merger	No	$p_{\text{features}} > 0.8$, $p_{\text{odd}<0.1}$, $p_{\text{merger}} < 0.1$	12
	Int	$p_{\text{odd}} > 0.5$, $0.1 < p_{\text{merger}} < 0.4$	12
	Yes	$p_{\text{odd}} > 0.5$, $p_{\text{merger}} > 0.4$	12
Edge-on	Yes	$p_{\text{edgeon}} > 0.8$, $p_{\text{features}} > 0.5$	12
	Int	$0.4 < p_{\text{edgeon}} < 0.8$, $p_{\text{features}} > 0.5$	12
	No	$p_{\text{edgeon}} < 0.2$, $p_{\text{features}} > 0.5$	12
Bar	No	$p_{\text{bar}} < 0.1$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} < 0.2$	24
	Int	$0.2 < p_{\text{bar}} < 0.4$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} < 0.2$	24
	Yes	$p_{\text{bar}} > 0.8$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} < 0.2$	24
Visible spiral	No	$p_{\text{spiral}} < 0.2$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} < 0.2$, $p_{\text{bar}} < 0.1$	12
	Int	$0.2 < p_{\text{spiral}} < 0.8$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} < 0.2$, $p_{\text{bar}} < 0.1$	12
	Yes	$p_{\text{spiral}} > 0.8$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} < 0.2$, $p_{\text{bar}} < 0.1$	12
Oblique bulge size	No	$p_{\text{nobulge}>0.6}$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} < 0.5$, $p_{\text{bar}} < 0.2$	12
	Int	$p_{\text{justnoticeable}} > 0.6$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} < 0.5$, $p_{\text{bar}} < 0.2$	12
	Yes	$p_{\text{obvious/dominant}} > 0.5$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} < 0.5$, $p_{\text{bar}} < 0.2$	12
Edge-on bulge shape	Round	$p_{\text{rounded}} > 0.5$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} > 0.5$	12
	Boxy	$p_{\text{boxy}} > 0.4$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} > 0.2$	12
	Non	$p_{\text{nobulge}} > 0.5$, $p_{\text{features}} > 0.5$, $p_{\text{edgeon}} > 0.5$	12
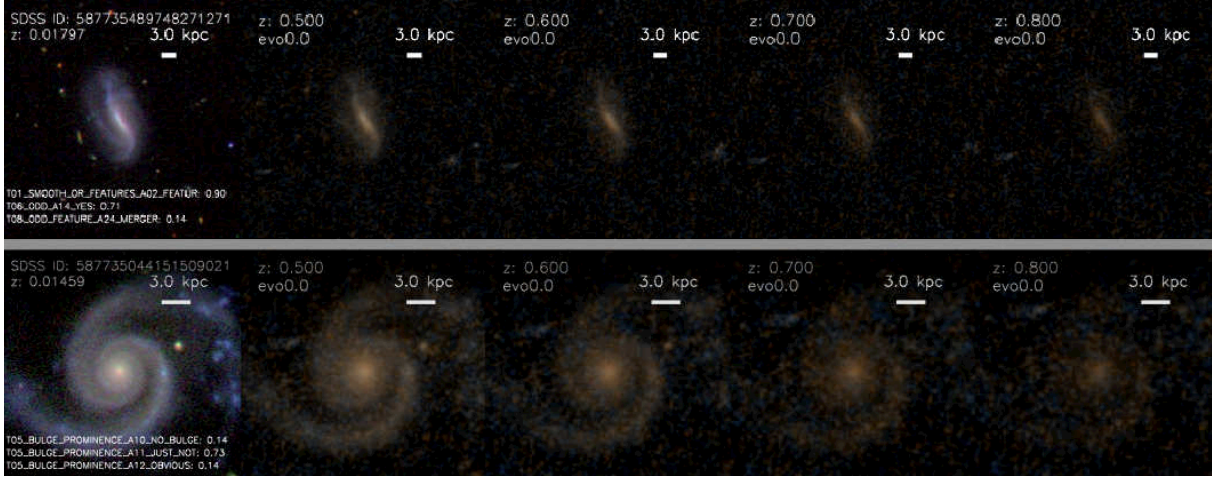
Figure 2. Examples of two galaxies which have been run through the FERENGI code to produce simulated HST images. The value of p_{features} for each panel is (1) Top row: $p_{\text{features}} = 0.9, 0.625, 0.35, 0.35, 0.225$ and (2) Bottom row: $p_{\text{features}} = 1.00, 0.875, 0.875, 0.625, 0.375$.

be used to measure the calibration to those with data at the pivot redshift of $z = 0.3$, those with non-zero f_{features} at $z = 0.3$, and with a reasonable fit to the exponential model ($\Delta\chi^2 > 3.0$).

Figure 4 shows the results of the sample; the correction is a weak function of surface brightness overall. Higher-surface brightness galaxies have stronger average corrections, likely because these galaxies are more likely to have larger f_{features} values at high redshifts. Low surface brightness galaxies are more likely to begin low and remain low; the bounded nature of the dropoff (and Poissonian-like variance among the individual voters) means that the average magnitude of ζ will be less.

We fit the data in Figure 4 with a linear function such that:

$$\log_{10}(\hat{\zeta}) = \zeta_0 + \zeta_1 \times \mu, \qquad (2)$$

where $\hat{\zeta}$ is the correction factor that will be applied to each galaxy as a function of surface brightness. The best-fit parameters to the linear fit (from least-squares optimization) are $\zeta_0 = 0.1, \zeta_1 = 1.4$. To make the final debiased correction, we modify the simple exponential form of Equation 1 to bound the debiased vote fractions between f and 1:

$$f_{\text{debiased}} = 1 - (1 - f)e^{\frac{z - z_0}{\hat{\zeta}}}. \qquad (3)$$
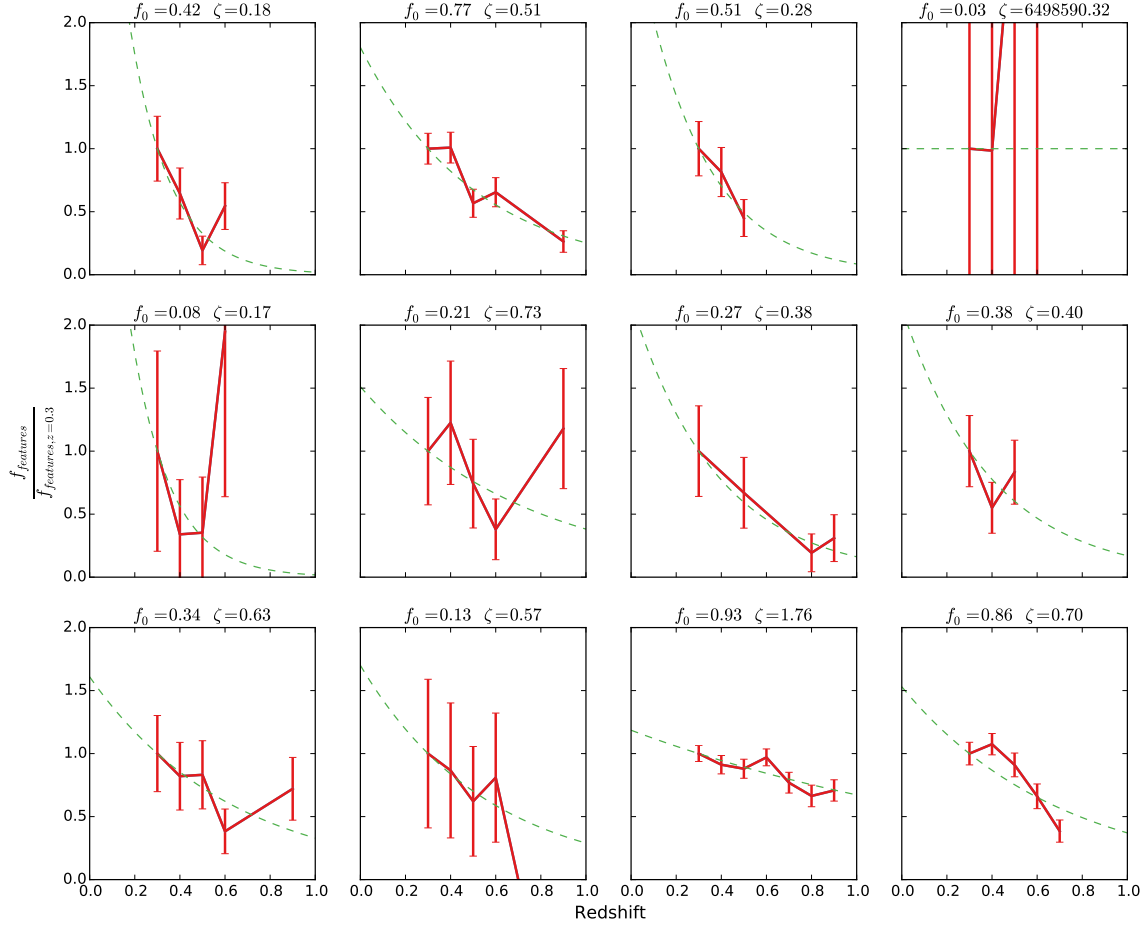
Figure 3. Behavior of the normalized, weighted vote fractions of features visible in a galaxy (f_{features}) as a function of redshift in the artifical FERENGI images. Galaxies are a random selection of images with $e = 0$ and at least three detectable images in redshift bins of $z \geq 0.3$. The measured vote fractions (red points) are fit with an exponential function (Equation 1); the best-fit parameters are given above each plot. Error bars are Poissonian, assuming a median of 40 votes per galaxy.

3.3 Results of Zeta Approach (Mel stuff)

In Figure 5 we examine the change in p_{features} for the FER-ENGI images relative to their lowest simulated redshift ($z_{\mathrm{sim}} = 0.3$). For each simulated redshift value z, and at fixed surface brightness μ, we plot $p_{\mathrm{features},z}$, the value measured at that simulated redshift, vs $p_{\mathrm{features},z=0.3}$, the value measured for the same galaxy imaged at $z = 0.3$. Only images corresponding to galaxies which also had FERENGI images at $z = 0.3$ were included in this plot, which include 4031 out of the total 6624.

Our objective is to use these data to predict, for a galaxy with a measured $p_{\mathrm{features},z}$ value, what its p_{features} value would have been if it had been viewed at $z = 0.3$. This predicted value is defined as the debiased vote fraction $p_{\mathrm{features,debiased}}$, and is calculated by applying a correction to the measured value of p_{features}, determined by the ζ function described in the previous section. We determine that a predicted value can be obtained so long as the relationship

between $p_{\mathrm{features},z}$ and $p_{\mathrm{features},z=0.3}$ is single-valued; that is, for a given $p_{\mathrm{features},z}$, there is one corresponding value of p_{features} at $z = 0.3$.

In Figure 5 we see that the relationship between $p_{\mathrm{features},z}$ and $p_{\mathrm{features},z=0.3}$ is *not* always single valued; hence we do not find it appropriate to correct galaxies that lie in these regions of surface brightness/redshift/p_{features} space. Our criteria for determining whether a region of this space is correctable is as follows: For each surface brightness and redshift bin, we examine the distribution of $p_{\mathrm{features},z=0.3}$ for fixed $p_{\mathrm{features},z}$ in bins of $\Delta p_{\mathrm{features},z} = 0.1$. These sub-bins are marked by the horizontal lines within each larger, surface brightness/redshift bin. First, we require at least 5 points in each sub-bin to determine with confidence whether that region is correctable or uncorrectable. For the region to be considered correctable, we require that the inner two quartiles of the distribution of $p_{\mathrm{features},z=0.3}$ within these sub-bins span a range less than 0.2. These regions are coloured pur-
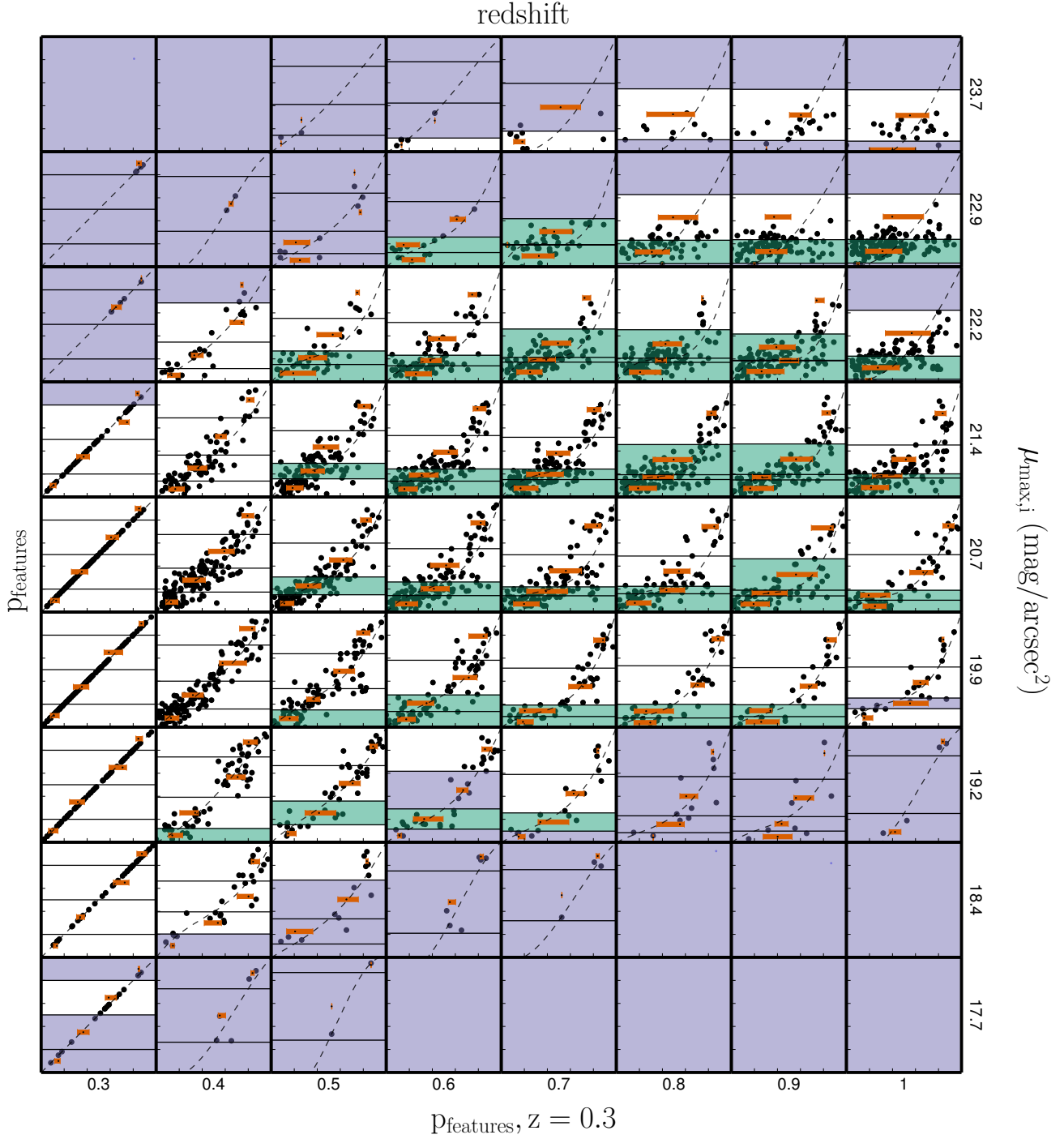
Figure 5. a caption. Add scale in p_{features} somehow that doesn't clutter the plot even more.

ple, and the inner two quartile ranges are indicated by the black error bars within each sub-bin.If the spread is greater than 0.2, the region is considered uncorrectable; these are marked in green. Any region with fewer than 5 points is considered to have "not enough information (NEI)", and left uncoloured on the plot.
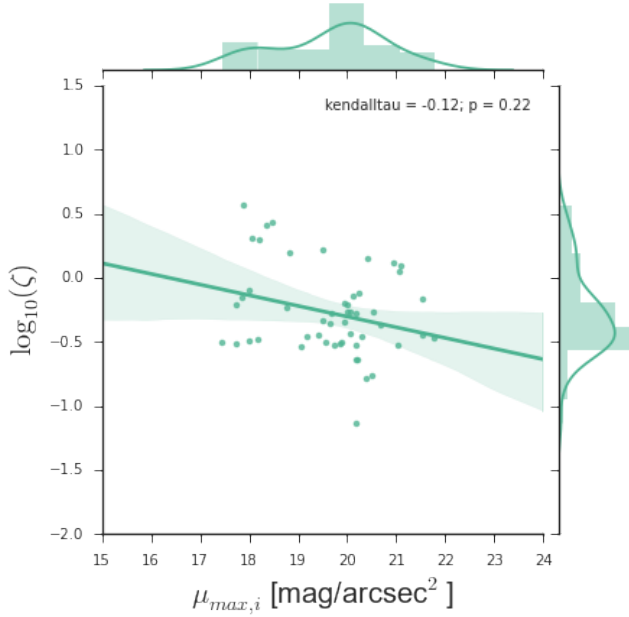
Stuff to do in this section:

• talk about where purple and green tend to be on plot,

and about how green areas are primarily ellipticals and disks with all the features washed out - can't tell them apart! hence, can't debias

• talk about where the Hubble sample falls in this space, reference table 4

• justify $N > 5$ and spread < 0.2 (or find a better way to choose criteria)

Figure 4. All fits for the vote fraction dropoff parameter ζ for f_{features} in the FERENGI galaxies as a function of surface brightness. This includes only the 51 galaxies with a reasonably bounded range on the dropoff ($-10 < \log(\zeta) < 10$) and sufficient points to fit the function.

Table 4. Debiasable Hubble Stuff. Doesn't include SDSS yet. Note: stuff without good redshift info automatically gets put in uncorrectable category - which is about 25,000 galaxies. Might increase correctable number if we can get these measurements.

Category	N	%
Correctable	7408	6 %
Uncorrectable	58526	50 %
NEI	52149	44 %
Total	118083	100 %

• check out corrections for correctable and NEI, show some sample images of corrected galaxies

• show some data for p_{bar}, determine or justify why we won't debiase them

3.4 Results of FERENGI Analysis

We show here the **preliminary** results of citizen scientist classifications of images of galaxies placed at artificial redshifts. Figure **??** shows the **preliminary** range of change of vote fractions for the change of p_{features} with redshift, for galaxies with different vote fraction levels, three ranges of surface-brightness levels and 7 evolutionary corrections (this last bit is indicated by the colour).

3.4.1 TODO LIST

We need to do:
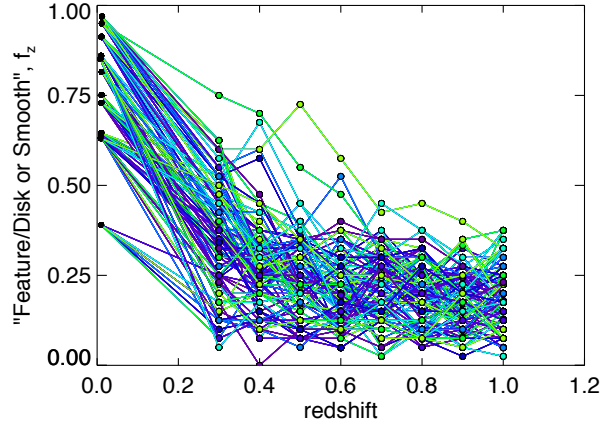
Figure 6. Preliminary results of the FERENGI redshifting exercise. WARNING NO USER WEIGHTING.... For a range of vote fraction levels with three surface-brightness levels and 7 evolutionary corrections each (the colours indicate evolutionary correction value with green being $e = 0$, we show the range of evolution of the vote fractions for featured vs. smooth with redshift.

• Calculate the magnitudes, surface brightnesses and sizes of the galaxies in the FERENGI images....

• Plot of magnitude distribution of galaxies in each of the four GZH subsamples with the magnitudes of our fake galaxies over plotted.

• Instructions of how to link the $z = 0$ p_X values for galaxies with a given size, magnitude (surface brightness) in the GZH images.

4 SUMMARY

Now people go and do science with these awesome Galaxy Zoo Hubble classifications.

REFERENCES

Bamford S. P. et al., 2009, MNRAS, 393, 1324
Barden M., Jahnke K., Häußler B., 2008, ApJS, 175, 105
Griffith R. L. et al., 2012, ApJS, 200, 9
Johnson L. C. et al., 2015, ApJ, 802, 127
Lintott C. J. et al., 2008, MNRAS, 389, 1179
Strauss M. A. et al., 2002, AJ, 124, 1810
Willett K. W. et al., 2013, MNRAS, 435, 2835
York D. G. et al., 2000, AJ, 120, 1579